

Achieving Secrecy Capacity of the Gaussian Wiretap Channel with Polar Lattices

Ling Liu, Yanfei Yan, and Cong Ling *Member, IEEE*

Abstract

In this work, an explicit wiretap coding scheme based on polar lattices is proposed to achieve the secrecy capacity of the additive white Gaussian noise (AWGN) wiretap channel. Firstly, polar lattices are used to construct secrecy-good lattices for the $\text{mod-}\Lambda_s$ Gaussian wiretap channel. Then we propose an explicit shaping scheme to remove this $\text{mod-}\Lambda_s$ front end and extend polar lattices to the genuine Gaussian wiretap channel. The shaping technique is based on the lattice Gaussian distribution, which leads to a binary asymmetric channel at each level for the multilevel lattice codes. By employing the asymmetric polar coding technique, we construct an AWGN-good lattice and a secrecy-good lattice with optimal shaping simultaneously. As a result, the encoding complexity for the sender and the decoding complexity for the legitimate receiver are both $O(N \log N \log(\log N))$. The proposed scheme is proven to be semantically secure.

I. INTRODUCTION

Wyner [1] introduced the wiretap channel model and showed that both reliability and confidentiality could be attained by coding without any key bits if the channel between the sender and the eavesdropper (wiretapper's channel W) is degraded with respect to the channel between the sender and the legitimate receiver (main channel V). The goal of wiretap coding is to design a coding scheme that makes it possible to communicate both reliably and securely between the sender and the legitimate receiver. Reliability is measured by the decoding error probability for the legitimate user, namely $\lim_{N \rightarrow \infty} \Pr\{\hat{M} \neq M\} = 0$, where N is the length of transmitted codeword, M is the confidential message and \hat{M} is its estimation. Secrecy is measured by the mutual information between M and the signal received by the eavesdropper $Z^{[N]}$. In this work, we will follow the strong secrecy condition proposed by Csiszár [2], i.e., $\lim_{N \rightarrow \infty} I(M; Z^{[N]}) = 0$, which is more widely accepted than the weak secrecy criterion $\lim_{N \rightarrow \infty} \frac{1}{N} I(M; Z^{[N]}) = 0$. In simple terms, the secrecy capacity is defined as the maximum achievable rate under both the reliability and strong secrecy conditions. When W is degraded with respect to V , the secrecy capacity is given by $C(V) - C(W)$ [3], where $C(\cdot)$ denotes the channel capacity.

This work was supported in part by FP7 project PHYLAWS (EU FP7-ICT 317562) and in part by the China Scholarship Council. This work was/will be presented at the IEEE Int. Symp. Inform. Theory (ISIT), Honolulu, USA, 2014 and the IEEE Inform. Theory Workshop, Jerusalem, ISRAEL, 2015.

Ling Liu, Yanfei Yan and Cong Ling are with the Department of Electrical and Electronic Engineering, Imperial College London, London, UK (e-mails: l.liu12@imperial.ac.uk, y.yan10@imperial.ac.uk, cling@ieee.org).

In the study of strong secrecy, plaintext messages are often assumed to be random and uniformly distributed. From a cryptographic point of view, it is crucial that the security does not rely on the distribution of the message. This issue can be resolved by using the standard notion of *semantic security* [4] which means that, asymptotically, it is impossible to estimate any function of the message better than to guess it without accessing $Z^{[N]}$ at all. The relation between strong secrecy and semantic security was recently revealed in [5], [6], namely, semantic security is equivalent to achieving strong secrecy for all distributions p_M of the plaintext messages:

$$\lim_{N \rightarrow \infty} \max_{p_M} I(M; Z^{[N]}) = 0. \quad (1)$$

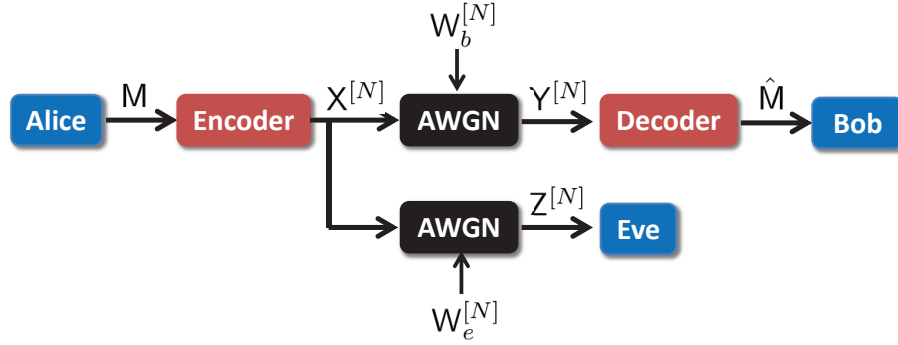


Fig. 1. The Gaussian wiretap channel.

In this work, we construct lattice codes for the Gaussian wiretap channel (GWC) which is shown in Fig. 1. The confidential message M drawn from the message set \mathcal{M} is encoded by the sender (Alice) into an N -dimensional codeword $X^{[N]}$. The outputs $Y^{[N]}$ and $Z^{[N]}$ received by the legitimate receiver (Bob) and the eavesdropper Eve are respectively given by

$$\begin{cases} Y^{[N]} = X^{[N]} + W_b^{[N]} \\ Z^{[N]} = X^{[N]} + W_e^{[N]}, \end{cases}$$

where $W_b^{[N]}$ and $W_e^{[N]}$ are N -dimensional Gaussian noise vectors with zero mean and variance σ_b^2 , σ_e^2 respectively. The channel input $X^{[N]}$ satisfies the power constraint P_s , i.e.,

$$\frac{1}{N} E[\|X^{[N]}\|^2] \leq P_s.$$

Polar codes [7] have shown their great potential in solving the wiretap coding problem. The polar coding scheme proposed in [8], combined with the block Markov coding technique [9], was proved to achieve the strong secrecy capacity when W and V are both binary-input symmetric channels, and W is degraded with respect to V . More recently, polar wiretap coding has been extended to general wiretap channels (not necessarily degraded or symmetric) in [10] and [11]. For continuous channels such as the GWC, there also has been notable progress in wiretap lattice coding. On the theoretical aspect, the existence of lattice codes achieving the secrecy capacity to within $\frac{1}{2}$ nat under

the strong secrecy as well as semantic security criterion was demonstrated in [6]. On the practical aspect, wiretap lattice codes were proposed in [12] and [13] to maximize the eavesdropper's decoding error probability.

A. Our contribution

Polar lattices, the counterpart of polar codes in the Euclidean space, have already been proved to be additive white Gaussian noise (AWGN)-good [14] and further to achieve the AWGN channel capacity with lattice Gaussian shaping [15]¹. Motivated by [8], we will propose polar lattices to achieve both strong secrecy and reliability over the mod- Λ_s GWC. Conceptually, this polar lattice structure can be regarded as a secrecy-good lattice Λ_e nested within an AWGN-good lattice Λ_b ($\Lambda_e \subset \Lambda_b$). Further, we will propose a Gaussian shaping scheme over Λ_b and Λ_e , using the multilevel asymmetric polar coding technique. As a result, we will accomplish the design of an explicit lattice coding scheme which achieves the secrecy capacity of the GWC. The novel technical contribution of this paper is two-fold:

- The construction of secrecy-good polar lattices for the mod- Λ_s GWC and the proof of their secrecy capacity-achieving. This is an extension of the binary symmetric wiretap coding [8] to the multilevel coding scenario, and can also be considered as the construction of secrecy-good polar lattices for the GWC without the power constraint. The construction for the mod- Λ_s GWC provides considerable insight into wiretap coding for the genuine GWC, without deviating to the technicality of Gaussian shaping. This work is also of independent interest to other problems of information theoretic security, e.g., secret key generation from Gaussian sources [19].
- The Gaussian shaping applied to the secrecy-good polar lattice, which follows the footpath of [15]. The resultant coding scheme is proved to achieve the secrecy capacity of the GWC. This coding scheme is further proved to be semantically secure. The idea follows the conception of [6], where lattice Gaussian sampling was employed to obtain semantic security. It is worth mentioning that our proposed coding scheme is not only a practical implementation of the secure random lattice coding in [6], but also an improvement in the sense that we successfully remove the constant $\frac{1}{2}$ -nat gap to the secrecy capacity.²

B. Comparison with the extractor-based approach

Invertible randomness extractors were introduced into wiretap coding in [5], [20], [21]. The key idea is that an extractor is used to convert a capacity-achieving code with rate close to $C(V)$ for the main channel into a wiretap code with the rate close to $C(V) - C(W)$. Later, this coding scheme was extended to the GWC in [22]. Besides, channel resolvability [23] was proposed as a tool for wiretap codes. An interesting connection between the resolvability and the extractor was revealed in [24].

¹Please refer to [16]–[18] for other methods of achieving the AWGN channel capacity.

²The $\frac{1}{2}$ -nat gap in [6] was due to a requirement on the flatness factor of the secrecy-good lattice. In this paper, we employ mutual information, rather than via the flatness factor, to directly bound information leakage, thereby removing that requirement of the secrecy-good lattice.

The proposed approach and the one based on invertible extractors have their respective advantages. The extractor-based approach is modular, i.e., the error-correction code and extractor are realized separately; it is possible to harness the results of invertible extractors in literature. The advantage of our lattice-based scheme is that the wiretap code designed for Eve is nested within the capacity-achieving code designed for Bob, which represents an integrated approach. More importantly, lattice codes are attractive for emerging applications in network information theory thanks to their useful structures [16], [25]; thus the proposed scheme may fit better with this landscape when security is a concern [26].

C. Outline of the paper

The paper is organized as follows: Section II presents some preliminaries of lattice codes. In Section III we construct secrecy-good polar lattices for the mod- Λ_s GWC, using the binary symmetric polar wiretap coding and multilevel lattice structure [27]. The original polar wiretap code in [8] is slightly modified to be compatible to the following shaping operation. In Section IV, we show how to implement the discrete Gaussian shaping over the polar lattice to remove the mod- Λ_s front end, using the polar coding technique for asymmetric channels. Then we prove that our wiretap lattice coding achieves the secrecy capacity with shaping. Furthermore, the strong secrecy is extended to semantic security. Finally, we discuss the relationship between the lattice constructions with and without shaping in Section V.

D. Notations

All random variables (RVs) will be denoted by capital letters. Let P_X denote the probability distribution of a RV X taking values x in a set \mathcal{X} and let $H(X)$ denote its entropy. For multilevel coding, we denote by X_ℓ a RV X at level ℓ . The i -th realization of X_ℓ is denoted by x_ℓ^i . We also use the notation $x_\ell^{i:j}$ as a shorthand for a vector $(x_\ell^i, \dots, x_\ell^j)$, which is a realization of RVs $X_\ell^{i:j} = (X_\ell^i, \dots, X_\ell^j)$. Similarly, $x_{\ell:j}^i$ will denote the realization of the i -th RVs from level ℓ to level j , i.e., of $X_{\ell:j}^i = (X_\ell^i, \dots, X_j^i)$. For a set \mathcal{I} , \mathcal{I}^c denotes its complement set, and $|\mathcal{I}|$ represents its cardinality. For an integer N , $[N]$ will be used to denote the set of all integers from 1 to N . W and \tilde{W} will be used to denote a binary memoryless asymmetric (BMA) channel and a binary memoryless symmetric (BMS) channel respectively. Following the notation of [7], we denote N independent uses of channel W by W^N . By channel combining and splitting, we get the combined channel W_N and the i -th subchannel $W_N^{(i)}$. Specifically, for a channel W_ℓ at level ℓ , W_ℓ^N , $W_{\ell,N}$ and $W_\ell^{(i,N)}$ are used to denote its N independent expansion, the combined channel and the i -th subchannel after polarization. $\mathbb{1}(\cdot)$ denotes the indicator function. Throughout this paper, we use the binary logarithm, denoted by \log , and information is measured in bits.

II. PRELIMINARIES OF LATTICE CODES

A. Definitions

A lattice is a discrete subgroup of \mathbb{R}^n which can be described by

$$\Lambda = \{\lambda = Bx : x \in \mathbb{Z}^n\},$$

where B is the n -by- n lattice generator matrix and we always assume that it has full rank in this paper.

For a vector $x \in \mathbb{R}^n$, the nearest-neighbor quantizer associated with Λ is $Q_\Lambda(x) = \arg \min_{\lambda \in \Lambda} \|\lambda - x\|$. We define the modulo lattice operation by $x \bmod \Lambda \triangleq x - Q_\Lambda(x)$. The Voronoi region of Λ , defined by $\mathcal{V}(\Lambda) = \{x : Q_\Lambda(x) = 0\}$, specifies the nearest-neighbor decoding region. The Voronoi cell is one example of fundamental region of the lattice. A measurable set $\mathcal{R}(\Lambda) \subset \mathbb{R}^n$ is a fundamental region of the lattice Λ if $\cup_{\lambda \in \Lambda} (\mathcal{R}(\Lambda) + \lambda) = \mathbb{R}^n$ and if $(\mathcal{R}(\Lambda) + \lambda) \cap (\mathcal{R}(\Lambda) + \lambda')$ has measure 0 for any $\lambda \neq \lambda'$ in Λ . The volume of a fundamental region is equal to that of the Voronoi region $\mathcal{V}(\Lambda)$, which is given by $\text{Vol}(\Lambda) = |\det(B)|$.

The theta series of Λ (see, e.g., [28, p.70]) is defined as

$$\Theta_\Lambda(\tau) = \sum_{\lambda \in \Lambda} e^{-\pi\tau\|\lambda\|^2}, \quad \tau > 0.$$

In this paper, to satisfy the reliability condition for Bob, we are mostly concerned with the block error probability $P_e(\Lambda, \sigma^2)$ of lattice decoding. It is the probability $\Pr\{x \notin \mathcal{V}(\Lambda)\}$ that an n -dimensional independent and identically distributed (i.i.d.) Gaussian noise vector x with zero mean and variance σ^2 per dimension falls outside the Voronoi region $\mathcal{V}(\Lambda)$. For an n -dimensional lattice Λ , define the volume-to-noise ratio (VNR) of Λ by

$$\gamma_\Lambda(\sigma) \triangleq \frac{\text{Vol}(\Lambda)^{\frac{2}{n}}}{\sigma^2}.$$

Then we introduce the notion of lattices which are good for the AWGN channel without power constraint.

Definition 1 (AWGN-good lattices): A sequence of lattices Λ_b of increasing dimension n is AWGN-good if, for any fixed $P_e(\Lambda_b, \sigma^2) \in (0, 1)$,

$$\lim_{n \rightarrow \infty} \gamma_{\Lambda_b}(\sigma) = 2\pi e$$

and if, for a fixed VNR greater than $2\pi e$, $P_e(\Lambda_b, \sigma^2)$ goes to 0 as $n \rightarrow \infty$.

It is worth mentioning here that we do not insist on exponentially vanishing error probabilities, unlike Poltyrev's original treatment of good lattices for coding over the AWGN channel [29]. This is because a sub-exponential or polynomial decay of the error probability is often good enough.

B. Flatness Factor and Lattice Gaussian Distribution

For $\sigma > 0$ and $c \in \mathbb{R}^n$, the Gaussian distribution of mean c and variance σ^2 is defined as

$$f_{\sigma, c}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{\|x - c\|^2}{2\sigma^2}},$$

for all $x \in \mathbb{R}^n$. For convenience, let $f_\sigma(x) = f_{\sigma, 0}(x)$.

Given lattice Λ , we define the Λ -periodic function

$$f_{\sigma, \Lambda}(x) = \sum_{\lambda \in \Lambda} f_{\sigma, \lambda}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \sum_{\lambda \in \Lambda} e^{-\frac{\|x - \lambda\|^2}{2\sigma^2}},$$

for $x \in \mathbb{R}^n$.

The flatness factor is defined for a lattice Λ as [6]

$$\epsilon_\Lambda(\sigma) \triangleq \max_{x \in \mathcal{R}(\Lambda)} |\text{Vol}(\Lambda) f_{\sigma, \Lambda}(x) - 1|.$$

It can be interpreted as the maximum variation of $f_{\sigma,\Lambda}(x)$ from the uniform distribution over $\mathcal{R}(\Lambda)$. The flatness factor can be calculated using the theta series [6]:

$$\epsilon_{\Lambda}(\sigma) = \left(\frac{\gamma_{\Lambda}(\sigma)}{2\pi} \right)^{\frac{n}{2}} \Theta_{\Lambda} \left(\frac{1}{2\pi\sigma^2} \right) - 1.$$

We define the *discrete Gaussian distribution* over Λ centered at $c \in \mathbb{R}^n$ as the following discrete distribution taking values in $\lambda \in \Lambda$:

$$D_{\Lambda,\sigma,c}(\lambda) = \frac{f_{\sigma,c}(\lambda)}{f_{\sigma,c}(\Lambda)}, \quad \forall \lambda \in \Lambda,$$

where $f_{\sigma,c}(\Lambda) \triangleq \sum_{\lambda \in \Lambda} f_{\sigma,c}(\lambda) = f_{\sigma,\Lambda}(c)$. Again for convenience, we write $D_{\Lambda,\sigma} = D_{\Lambda,\sigma,0}$.

It is also useful to define the discrete Gaussian distribution over a coset of Λ , i.e., the shifted lattice $\Lambda - c$:

$$D_{\Lambda-c,\sigma}(\lambda - c) = \frac{f_{\sigma}(\lambda - c)}{f_{\sigma,c}(\Lambda)}, \quad \forall \lambda \in \Lambda.$$

Note the relation $D_{\Lambda-c,\sigma}(\lambda - c) = D_{\Lambda,\sigma,c}(\lambda)$, namely, they are a shifted version of each other.

Each component of a lattice point sampled from $D_{\Lambda-c,\sigma}$ has an average power always less than σ^2 by the following lemma.

Lemma 1 (Average power of lattice Gaussian [30]): Let $x = (x_1, x_2, \dots, x_n)^T \sim D_{\Lambda-c,\sigma}$. Then, for each $1 \leq i \leq n$,

$$E[x_i^2] \leq \sigma^2. \quad (2)$$

If the flatness factor is negligible, the discrete Gaussian distribution over a lattice preserves the capacity of the AWGN channel.

Theorem 1 (Mutual information of discrete Gaussian distribution [30]): Consider an AWGN channel $Y = X + E$ where the input constellation X has a discrete Gaussian distribution $D_{\Lambda-c,\sigma_s}$ for arbitrary $c \in \mathbb{R}^n$, and where the variance of the noise E is σ^2 . Let the average signal power be P_s so that $\text{SNR} = P_s/\sigma^2$, and let $\tilde{\sigma} \triangleq \frac{\sigma_s\sigma}{\sqrt{\sigma_s^2 + \sigma^2}}$. Then, if $\varepsilon = \epsilon_{\Lambda}(\tilde{\sigma}) < \frac{1}{2}$ and $\frac{\pi\varepsilon_t}{1-\varepsilon_t} \leq \varepsilon$ where

$$\varepsilon_t \triangleq \begin{cases} \epsilon_{\Lambda} \left(\sigma_s / \sqrt{\frac{\pi}{\pi-t}} \right), & t \geq 1/e \\ (t^{-4} + 1) \epsilon_{\Lambda} \left(\sigma_s / \sqrt{\frac{\pi}{\pi-t}} \right), & 0 < t < 1/e \end{cases}$$

the discrete Gaussian constellation results in mutual information

$$I_D \geq \frac{1}{2} \log(1 + \text{SNR}) - \frac{5\varepsilon}{n} \quad (3)$$

per channel use.

A lattice Λ or its coset $\Lambda - c$ with a discrete Gaussian distribution is referred to as a *good constellation* for the AWGN channel if $\epsilon_{\Lambda}(\tilde{\sigma})$ is negligible [30]. It is further proved in [30] that the channel capacity is achieved with Gaussian shaping over an AWGN-good lattice and minimum mean square error (MMSE) lattice decoding. Following Theorem 1, it has been shown in [15] that an AWGN-good polar lattice shaped according to the discrete Gaussian distribution achieves the AWGN channel capacity with sub-exponentially vanishing error probability, which means

that an explicit polar lattice satisfying the power constraint and the reliability condition for Bob is already in hand. Therefore, the next section will focus on the construction of the secrecy-good polar lattice.

III. SECRECY-GOOD POLAR LATTICES FOR THE MOD- Λ_s GWC

A. Polar codes: brief review

We firstly recall some basics of polar codes. Let \tilde{W} be a BMS channel with uniformly distributed input $X \in \mathcal{X} = \{0, 1\}$ and output $Y \in \mathcal{Y}$. The input distribution and transition probability of \tilde{W} are denoted by P_X and $P_{Y|X}$ respectively. Let $X^{[N]}$ and $Y^{[N]}$ be the input and output vector of N independent uses of \tilde{W} . Suppose $N = 2^m$ for some integer $m \geq 1$, the channel polarization is resulted from the transform $U^{[N]} = X^{[N]}G_N$ where $G_N = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes m}$ is the generator matrix and \otimes denotes the Kronecker product. Then we get an N -dimensional combined channel \tilde{W}_N from $U^{[N]}$ to $Y^{[N]}$. For each $i \in [N]$, given the previous bits $U^{1:i-1}$, the channel $\tilde{W}_N^{(i)}$ seen by each bit U^i is called the i -th subchannel channel after the channel splitting process [7], and the transition probability of $\tilde{W}_N^{(i)}$ is given by

$$\tilde{W}_N^{(i)}(y^{[N]}, u^{1:i-1} | u^i) = \sum_{u^{i+1:N} \in \mathcal{X}^{N-i}} \frac{1}{2^{N-1}} \tilde{W}_N(y^{[N]} | u^{[N]}),$$

where $u^{[N]}$ and $y^{[N]}$ are the realizations of $U^{[N]}$ and $Y^{[N]}$, respectively. Arıkan proved that $\tilde{W}_N^{(i)}$ is also a BMS channel and it becomes either an almost error-free channel or a completely useless channel as N grows. According to [7], the goodness of a BMS channel can be estimated by its associate Bhattacharyya parameter, which is defined as follows.

Definition 2 (Bhattacharyya parameter of BMS channels): Let \tilde{W} be a BMS channel with transition probability $P_{Y|X}$, the symmetric Bhattacharyya parameter $\tilde{Z} \in [0, 1]$ is defined as

$$\tilde{Z}(\tilde{W}) \triangleq \sum_y \sqrt{P_{Y|X}(y|0)P_{Y|X}(y|1)}.$$

It was further shown in [31], [32] that for any $\beta < \frac{1}{2}$,

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{1}{N} \left| \{i : \tilde{Z}(\tilde{W}_N^{(i)}) < 2^{-N^\beta}\} \right| &= I(\tilde{W}) \\ \lim_{m \rightarrow \infty} \frac{1}{N} \left| \{i : \tilde{Z}(\tilde{W}_N^{(i)}) > 1 - 2^{-N^\beta}\} \right| &= 1 - I(\tilde{W}), \end{aligned}$$

which means the proportion of such roughly error-free subchannels (with negligible Bhattacharyya parameters) approaches the channel capacity $I(\tilde{W})$. The set of the indices of all those almost error-free subchannels is usually called the information set \mathcal{I} and its complementary is called the frozen set \mathcal{F} . Consequently, the construction of capacity-achieving polar codes is simply to identify the indices in the information set \mathcal{I} . However, for a general BMS channel other than binary erasure channel, the complexity of the exact computation for $\tilde{Z}(\tilde{W}_N^{(i)})$ appears to be exponential in the block length N . An efficient estimation method for $\tilde{Z}(\tilde{W}_N^{(i)})$ was proposed in [33], using the idea of channel upgrading and degrading. It was shown that with a sufficient number of quantization levels, the approximation error is negligible even if \tilde{W} has continuous output, and the involved computational complexity is acceptable.

In [7], a bit-wised decoding method called successive cancellation (SC) decoding was proposed to show that polar codes are able to achieve channel capacity with vanishing error probability. This decoding method has complexity $O(N \log N)$, and the error probability is given by $P_e^{SC} \leq \sum_{i \in \mathcal{I}} \tilde{Z}(\tilde{W}_N^{(i)})$.

B. Polar codes for the binary symmetric wiretap channel

Now we revisit the construction of polar codes for the binary symmetric wiretap channel. We use \tilde{V} and \tilde{W} to denote the symmetric main channel between Alice and Bob and the symmetric wiretap channel between Alice and Eve, respectively. Both \tilde{V} and \tilde{W} have binary input X and \tilde{W} is degraded with respect to \tilde{V} . Let Y and Z denote the output of \tilde{V} and \tilde{W} . After the channel combination and splitting of N independent uses of the \tilde{V} and \tilde{W} by the polarization transform $U^{[N]} = X^{[N]} G_N$, we define the sets of reliability-good indices for Bob and information-poor indices for Eve as

$$\begin{aligned} \mathcal{G}(\tilde{V}) &= \{i : \tilde{Z}(\tilde{V}_N^{(i)}) \leq 2^{-N^\beta}\}, \\ \mathcal{N}(\tilde{W}) &= \{i : \tilde{Z}(\tilde{W}_N^{(i)}) \geq 1 - 2^{-N^\beta}\}, \end{aligned} \quad (4)$$

where $0 < \beta < 0.5$ and $\tilde{V}_N^{(i)}$ ($\tilde{W}_N^{(i)}$) is the i -th subchannel of the main channel (wiretapper's channel) after polarization transform.

Note that in the seminal paper [8] of polar wiretap coding, the information-poor set $\mathcal{N}(\tilde{W})$ was defined as $\{i : I(\tilde{W}^{(i,N)}) \leq 2^{-N^\beta}\}$. In contrast, our criterion here is based on the Bhattacharyya parameter³. This slight modification will bring us much convenience when lattice shaping is involved in Sect. IV. The following lemma shows that the modified criterion is similar to the original one in the sense that the mutual information of the subchannels with indices in $\mathcal{N}(\tilde{W})$ can still be bounded in the same form.

Lemma 2: Let $\tilde{W}_N^{(i)}$ be the i -th subchannel after the polarization transform on independent N uses of a BMS channel \tilde{W} . If $\tilde{Z}(\tilde{W}_N^{(i)}) \geq 1 - 2^{-N^\beta}$, the mutual information of the i -th subchannel can be upper-bounded as

$$I(\tilde{W}_N^{(i)}) \leq 2^{-N^{\beta'}}, 0 < \beta' < \beta < 0.5,$$

for sufficiently large N .

Proof: When \tilde{W} is symmetric, $\tilde{W}_N^{(i)}$ is symmetric as well. By [7, Proposition 1], we have

$$\begin{aligned} I(\tilde{W}_N^{(i)}) &\leq \sqrt{1 - \tilde{Z}(\tilde{W}_N^{(i)})^2} \\ &\leq \sqrt{2 \cdot 2^{-N^\beta}} \leq 2^{-N^{\beta'}}, \end{aligned}$$

where the last inequality holds for sufficiently large N . \square

Since the mutual information of subchannels in $\mathcal{N}(\tilde{W})$ can be upper-bounded in the same form, it is not difficult to understand that strong secrecy can be achieved using the index partition proposed in [8]. Similarly, we divide the index set $[N]$ into the following four sets:

$$\begin{aligned} \mathcal{A} &= \mathcal{G}(\tilde{V}) \cap \mathcal{N}(\tilde{W}), \quad \mathcal{B} = \mathcal{G}(\tilde{V}) \cap \mathcal{N}(\tilde{W})^c \\ \mathcal{C} &= \mathcal{G}(\tilde{V})^c \cap \mathcal{N}(\tilde{W}), \quad \mathcal{D} = \mathcal{G}(\tilde{V})^c \cap \mathcal{N}(\tilde{W})^c. \end{aligned} \quad (5)$$

³This idea has already been used in [8] to prove that polar wiretap coding scheme is secrecy capacity-achieving.

Clearly, $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C} \cup \mathcal{D} = [N]$. Then we assign set \mathcal{A} with message bits M , set \mathcal{B} with random bits R , set \mathcal{C} with frozen bits F which are known to both Bob and Eve prior to transmission, and set \mathcal{D} with random bits R . The next lemma shows that this assignment achieves strong secrecy. We note that this proof is similar to that in [9] and it is given in Appendix A.

Lemma 3: According to the partitions of the index set shown in (5), if we assign the four sets as follows

$$\begin{aligned} \mathcal{A} &\leftarrow M, & \mathcal{B} &\leftarrow R, \\ \mathcal{C} &\leftarrow F, & \mathcal{D} &\leftarrow R, \end{aligned} \tag{6}$$

the information leakage $I(M; Z^{[N]})$ can be upper-bounded as

$$I(M; Z^{[N]}) \leq N \cdot 2^{-N^{\beta'}}, 0 < \beta' < 0.5. \tag{7}$$

With regard to the secrecy rate, we show that the modified polar coding scheme can also achieve the secrecy capacity.

Lemma 4: Let $C(\tilde{V})$ and $C(\tilde{W})$ denote the channel capacity of the main channel \tilde{V} and wiretap channel \tilde{W} respectively. Since \tilde{W} is degraded with respect to \tilde{V} , the secrecy capacity, which is given by $C(\tilde{V}) - C(\tilde{W})$, is achievable using the modified wiretap coding scheme, i.e.,

$$\lim_{N \rightarrow \infty} |\mathcal{G}(\tilde{V}) \cap \mathcal{N}(\tilde{W})|/N = C(\tilde{V}) - C(\tilde{W}).$$

Proof: See Appendix B. □

We can also observe that the proportion of the problematic set \mathcal{D} is arbitrarily small when N is sufficiently large. This is because set \mathcal{D} is a subset of the unpolarized set $\{i : 2^{-N^\beta} < \tilde{Z}(\tilde{V}_N^{(i)}) < 1 - 2^{-N^\beta}\}$. As has been shown in [8], the reliability condition cannot be fulfilled with SC decoding due to the existence of \mathcal{D} . Fortunately, we can use the blocking technique proposed in [9] to achieve reliability and strong secrecy simultaneously. More details of this blocking technique will be discussed in Section III-D and Section IV-E.

C. Secrecy-good polar lattices

A sublattice $\Lambda' \subset \Lambda$ induces a partition (denoted by Λ/Λ') of Λ into equivalence classes modulo Λ' . The order of the partition is denoted by $|\Lambda/\Lambda'|$, which is equal to the number of cosets. If $|\Lambda/\Lambda'| = 2$, we call this a binary partition. Let $\Lambda/\Lambda_1/\dots/\Lambda_{r-1}/\Lambda'$ for $r \geq 1$ be an n -dimensional lattice partition chain. For each partition $\Lambda_{\ell-1}/\Lambda_\ell$ ($1 \leq \ell \leq r$ with convention $\Lambda_0 = \Lambda$ and $\Lambda_r = \Lambda'$) a code C_ℓ over $\Lambda_{\ell-1}/\Lambda_\ell$ selects a sequence of representatives a_ℓ for the cosets of Λ_ℓ . Consequently, if each partition is binary, the code C_ℓ is a binary code.

Polar lattices are constructed by ‘‘Construction D’’ [28, p.232] using a set of nested polar codes $C_1 \subseteq C_2 \subseteq \dots \subseteq C_r$ [27]. Suppose C_ℓ has block length N and the number of information bits k_ℓ for $1 \leq \ell \leq r$. Choose a basis $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_N$ from the polar generator matrix G_N such that $\mathbf{g}_1, \dots, \mathbf{g}_{k_\ell}$ span C_ℓ . When the dimension $n = 1$, the lattice L admits the form [27]

$$L = \left\{ \sum_{\ell=1}^r 2^{\ell-1} \sum_{i=1}^{k_\ell} u_\ell^i \mathbf{g}_i + 2^r \mathbb{Z}^N \mid u_\ell^i \in \{0, 1\} \right\}, \tag{8}$$

where the addition is carried out in \mathbb{R}^N . The fundamental volume of a lattice obtained from this construction is given by

$$\text{Vol}(L) = 2^{-NR_C} \cdot \text{Vol}(\Lambda_r)^N,$$

where $R_C = \sum_{\ell=1}^r R_\ell = \frac{1}{N} \sum_{\ell=1}^r k_\ell$ denotes the sum rate of component codes. In this paper, we limit ourselves to the binary lattice partition chain and binary polar codes for simplicity.

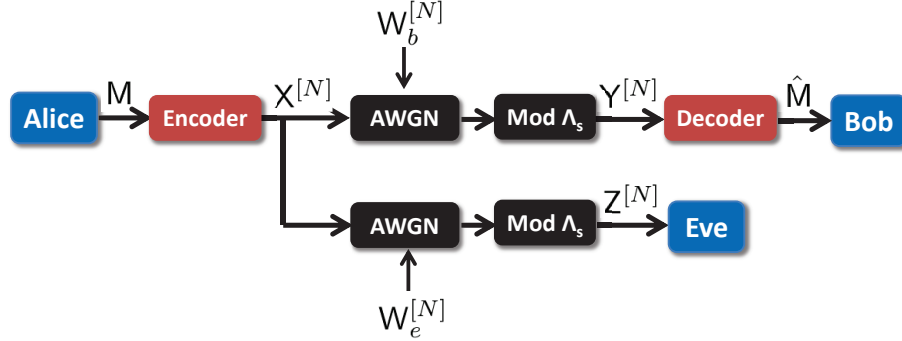


Fig. 2. The mod- Λ_s Gaussian wiretap channel.

Now we consider the construction of secrecy-good polar lattices over the mod- Λ_s GWC shown in Fig. 2. The difference between the mod- Λ_s GWC and the genuine GWC is the mod- Λ_s operation on the received signal of Bob and Eve. With some abuse of notation, the outputs $Y^{[N]}$ and $Z^{[N]}$ at Bob and Eve's ends respectively become

$$\begin{cases} Y^{[N]} = [X^{[N]} + W_b^{[N]}] \bmod \Lambda_s, \\ Z^{[N]} = [X^{[N]} + W_e^{[N]}] \bmod \Lambda_s. \end{cases}$$

The idea of wiretap lattice coding over the mod- Λ_s GWC [6] can be explained as follows. Let Λ_b and Λ_e be the AWGN-good lattice and secrecy-good lattice designed for Bob and Eve accordingly. Let $\Lambda_s \subset \Lambda_e \subset \Lambda_b$ be a nested chain of N -dimensional lattices in \mathbb{R}^N , where Λ_s is the shaping lattice. Note that the shaping lattice Λ_s here is employed primarily for the convenience of designing the secrecy-good lattice and secondarily for satisfying the power constraint. Consider a one-to-one mapping: $\mathcal{M} \rightarrow \Lambda_b/\Lambda_e$ which associates each message $m \in \mathcal{M}$ to a coset $\tilde{\lambda}_m \in \Lambda_b/\Lambda_e$. Alice selects a lattice point $\lambda \in \Lambda_e \cap \mathcal{V}(\Lambda_s)$ uniformly at random and transmits $X^{[N]} = \lambda + \lambda_m$, where λ_m is the coset representative of $\tilde{\lambda}_m$ in $\mathcal{V}(\Lambda_e)$. This scheme has been proved to achieve both reliability and semantic security in [6] by random lattice codes. We will make it explicit by constructing polar lattice codes in this section.

Let Λ_b and Λ_e be constructed from a binary partition chain $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda_r$, and assume $\Lambda_s \subset \Lambda_r^N$ such that $\Lambda_s \subset \Lambda_r^N \subset \Lambda_e \subset \Lambda_b^4$. Also, denote by $X_{1:r}^{[N]}$ the bits encoding Λ^N/Λ_r^N , which include all information bits

⁴This is always possible with sufficient power, since the power constraint is not our primary concern in this section.

for message M as a subset. We have that $[X^{[N]} + W_e^{[N]}] \bmod \Lambda_r^N$ is a sufficient statistic for $X_{1:r}^{[N]}$. This can be seen from [27, Lemma 8], rewritten as follows:

Lemma 5 (Sufficiency of mod- Λ output [27]): For a partition chain Λ/Λ' ($\Lambda' \subset \Lambda$), let the input of an AWGN channel be $X = A + B$, where $A \in \mathcal{R}(\Lambda)$ is a random variable, and B is uniformly distributed in $\Lambda \cap \mathcal{R}(\Lambda')$. Reduce the output Y first to $Y' = Y \bmod \Lambda'$ and then to $Y'' = Y' \bmod \Lambda$. Then the mod- Λ map is information-lossless, namely $I(A; Y') = I(A; Y'')$, which means that the output $Y'' = Y' \bmod \Lambda$ of mod- Λ map is a sufficient statistic for A .

In our context, we identify Λ with Λ_r^N and Λ' with Λ_s , respectively. Since the bits encoding Λ_r^N/Λ_s are uniformly distributed⁵, the mod- Λ_r^N operation is information-lossless in the sense that

$$I(X_{1:r}^{[N]}; Z^{[N]}) = I(X_{1:r}^{[N]}; [X^{[N]} + W_e^{[N]}] \bmod \Lambda_r^N).$$

As far as mutual information $I(X_{1:r}^{[N]}; Z^{[N]})$ is concerned, we can use the mod- Λ_r^N operator instead of the mod- Λ_s operator here. Under this condition, similarly to the multilevel lattice structure introduced in [27], the mod- Λ_s channel can be decomposed into a series of BMS channels according to the partition chain $\Lambda/\Lambda_1/\dots/\Lambda_{r-1}/\Lambda_r$. Therefore, the already mentioned polar coding technique for BMS channels can be employed. Moreover, the channel resulted from the lattice partition chain can be proved to be equivalent to that based on the chain rule of mutual information. Following this channel equivalence, we can construct an AWGN-good lattice Λ_b and a secrecy-good lattice Λ_e , using the wiretap coding technique (4) at each partition level.

A mod- Λ channel is a Gaussian channel with a modulo- Λ operator in the front end [27], [34]. The capacity of the mod- Λ channel is [27]

$$C(\Lambda, \sigma^2) = \log(\text{Vol}(\Lambda)) - h(\Lambda, \sigma^2), \quad (9)$$

where $h(\Lambda, \sigma^2)$ is the differential entropy of the Λ -aliased noise over $\mathcal{V}(\Lambda)$:

$$h(\Lambda, \sigma^2) = - \int_{\mathcal{V}(\Lambda)} f_{\sigma, \Lambda}(t) \log f_{\sigma, \Lambda}(t) dt.$$

The differential entropy is maximized to $\log(\text{Vol}(\Lambda))$ by the uniform distribution over $\mathcal{V}(\Lambda)$. The $\Lambda_{\ell-1}/\Lambda_\ell$ channel is defined as a mod- Λ_ℓ channel whose input is drawn from $\Lambda_{\ell-1} \cap \mathcal{V}(\Lambda_\ell)$. It is known that the $\Lambda_{\ell-1}/\Lambda_\ell$ channel is symmetric⁶, and the optimum input distribution is uniform [27]. Furthermore, the $\Lambda_{\ell-1}/\Lambda_\ell$ channel is binary if $|\Lambda_{\ell-1}/\Lambda_\ell| = 2$. The capacity of the $\Lambda_{\ell-1}/\Lambda_\ell$ channel for Gaussian noise of variance σ^2 is given by [27]

$$\begin{aligned} C(\Lambda_{\ell-1}/\Lambda_\ell, \sigma^2) &= C(\Lambda_\ell, \sigma^2) - C(\Lambda_{\ell-1}, \sigma^2) \\ &= h(\Lambda_{\ell-1}, \sigma^2) - h(\Lambda_\ell, \sigma^2) + \log(\text{Vol}(\Lambda_\ell)/\text{Vol}(\Lambda_{\ell-1})). \end{aligned}$$

The decomposition into a set of $\Lambda_{\ell-1}/\Lambda_\ell$ channels is used in [27] to construct AWGN-good lattices. Take the partition chain $\mathbb{Z}/2\mathbb{Z}/\dots/2^r\mathbb{Z}$ as an example. Given uniform input $X_{1:r}$, let \mathcal{K}_ℓ denote the coset indexed by $x_{1:\ell}$,

⁵In fact, all bits encoding Λ_e/Λ_s are uniformly distributed in wiretap coding.

⁶This is “regular” in the sense of Delsarte and Piret and symmetric in the sense of Gallager [27].

i.e., $\mathcal{K}_\ell = x_1 + \dots + 2^{\ell-1}x_\ell + 2^\ell\mathbb{Z}$. Given that $X_{1:\ell-1} = x_{1:\ell-1}$, the conditional probability distribution function (PDF) of this channel with binary input X_ℓ and output $\bar{Z} = Z \bmod \Lambda_\ell$ is

$$f_{\bar{Z}|X_\ell}(\bar{z}|x_\ell) = \frac{1}{\sqrt{2\pi}\sigma_e} \sum_{a \in \mathcal{K}_\ell(x_{1:\ell})} \exp\left(-\frac{1}{2\sigma_e^2}\|\bar{z} - a\|^2\right). \quad (10)$$

Since the previous input bits $x_{1:\ell-1}$ cause a shift on \mathcal{K}_ℓ and will be removed by the multistage decoder at level ℓ , the code can be designed according to the channel transition probability (10) with $x_{1:\ell-1} = 0$. Following the notation of [27], we use $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ and $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ to denote the $\Lambda_{\ell-1}/\Lambda_\ell$ channel for Bob and Eve respectively. The $\Lambda_{\ell-1}/\Lambda_\ell$ channel can also be used to construct secrecy-good lattices. In order to bound the information leakage of the wiretapper's channel, we firstly express $I(X_{1:r}; Z)$ according to the chain rule of mutual information as

$$I(X_{1:r}; Z) = I(X_1; Z) + I(X_2; Z|X_1) + \dots + I(X_r; Z|X_{1:r-1}). \quad (11)$$

This equation still holds if Z denotes the noisy signal after the mod- Λ_r operation, namely, $Z = [X + W_e] \bmod \Lambda_r$. We will adopt this notation in the rest of this subsection. We refer to the ℓ -th channel associated with mutual information $I(X_\ell; Z|X_{1:\ell-1})$ as the equivalent channel denoted by $W'(X_\ell; Z|X_{1:\ell-1})$, which is defined as the channel from X_ℓ to Z given the previous $X_{1:\ell-1}$. Then the transition probability distribution of $W'(X_\ell; Z|X_{1:\ell-1})$ is [27, Lemma 6]

$$\begin{aligned} f_{Z|X_\ell}(z|x_\ell) &= \frac{1}{\Pr(\mathcal{K}_\ell(x_{1:\ell}))} \sum_{a \in \mathcal{K}_\ell(x_{1:\ell})} \Pr(a) f_Z(z|a) \\ &= \frac{1}{|\Lambda_\ell/\Lambda_r|} \frac{1}{\sqrt{2\pi}\sigma_e} \sum_{a \in \mathcal{K}_\ell(x_{1:\ell})} \exp\left(-\frac{1}{2\sigma_e^2}\|z - a\|^2\right), \quad z \in \mathcal{V}(\Lambda_r). \end{aligned} \quad (12)$$

From (10) and (12), we can observe that the channel output likelihood ratio (LR) of the $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ channel is equal to that of the ℓ -th equivalent channel $W'(X_\ell; Z|X_{1:\ell-1})$. Then we have the following channel equivalence lemma.

Lemma 6: Consider a lattice L constructed by a binary lattice partition chain $\Lambda/\Lambda_1/\dots/\Lambda_{r-1}/\Lambda_r$. Constructing a polar code for the ℓ -th equivalent binary-input channel $W'(X_\ell; Z|X_{1:\ell-1})$ defined by the chain rule (11) is equivalent to constructing a polar code for the $\Lambda_{\ell-1}/\Lambda_\ell$ channel $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$.

Proof: See Appendix C. □

Note that another proof based on direct calculation of the mutual information and Bhattacharyya parameters of the subchannels can be found in [35].

Remark 1: Observe that if we define $V'(X_\ell; Y|X_{1:\ell-1})$ as the equivalent channel according to the chain rule expansion of $I(X; Y)$ for the main channel, the same result can be obtained between $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ and $V'(X_\ell; Y|X_{1:\ell-1})$. Moreover, this lemma also holds without the mod- Λ_s front-end, i.e., without power constraint. The construction of AWGN-good polar lattices was given in [15], where nested polar codes were constructed based on a set of $\Lambda_{\ell-1}/\Lambda_\ell$ channels. We note that the $\Lambda_{\ell-1}/\Lambda_\ell$ channel is degraded with respect to the $\Lambda_\ell/\Lambda_{\ell+1}$ channel [15, Lemma 3].

Now it is ready to introduce the polar lattice construction for the mod- Λ_s GWC shown in Fig. 3. A polar lattice L is constructed by a series of nested polar codes $C_1(N, k_1) \subseteq C_2(N, k_2) \subseteq \dots \subseteq C_r(N, k_r)$ and a binary

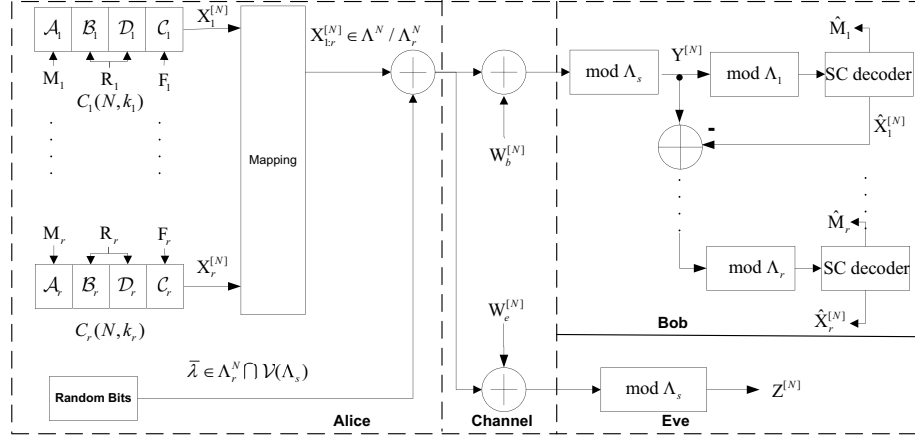


Fig. 3. The multilevel lattice coding system over the $\text{mod-}\Lambda_s$ Gaussian wiretap channel.

lattice partition chain $\Lambda/\Lambda_1/\dots/\Lambda_r$. The block length of polar codes is N . Alice splits the message M into M_1, \dots, M_r . We follow the same rule (6) to assign bits in the component polar codes to achieve strong secrecy. Note that $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ is degraded with respect to $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ for $1 \leq \ell \leq r$ because $\sigma_b^2 \leq \sigma_e^2$. Treating $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ and $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ as the main channel and wiretapper's channel at each level and using the partition rule (5), we can get four sets $\mathcal{A}_\ell, \mathcal{B}_\ell, \mathcal{C}_\ell$ and \mathcal{D}_ℓ . Similarly, we assign the bits as follows

$$\begin{aligned} \mathcal{A}_\ell &\leftarrow M_\ell, \quad \mathcal{B}_\ell \leftarrow R_\ell, \\ \mathcal{C}_\ell &\leftarrow F_\ell, \quad \mathcal{D}_\ell \leftarrow R_\ell \end{aligned} \quad (13)$$

for each level ℓ , where M_ℓ, F_ℓ and R_ℓ represent message bits, frozen bits (could be set as all zeros) and random bits at level ℓ . Since the $\Lambda_{\ell-1}/\Lambda_\ell$ channel is degraded with respect to the $\Lambda_\ell/\Lambda_{\ell+1}$ channel, it is easy to obtain that $\mathcal{C}_\ell \supseteq \mathcal{C}_{\ell+1}$, which means $\mathcal{A}_\ell \cup \mathcal{B}_\ell \cup \mathcal{D}_\ell \subseteq \mathcal{A}_{\ell+1} \cup \mathcal{B}_{\ell+1} \cup \mathcal{D}_{\ell+1}$. This construction is clearly a lattice construction as polar codes constructed on each level are nested. We skip the proof of nested polar codes here. A similar proof can be found in [14].

As a result, the above multilevel construction yields an AWGN-good lattice Λ_b and a secrecy-good lattice Λ_e simultaneously⁷. More precisely, Λ_b is constructed from a set of nested polar codes $C_1(N, |\mathcal{A}_1| + |\mathcal{B}_1| + |\mathcal{D}_1|) \subseteq \dots \subseteq C_r(N, |\mathcal{A}_r| + |\mathcal{B}_r| + |\mathcal{D}_r|)$, while Λ_e is constructed from a set of nested polar codes $C_1(N, |\mathcal{B}_1| + |\mathcal{D}_1|) \subseteq \dots \subseteq C_r(N, |\mathcal{B}_r| + |\mathcal{D}_r|)$ and with the same lattice partition chain. Note that the random bits in set \mathcal{D}_ℓ should be shared to Bob to guarantee the AWGN-goodness of Λ_b . More details are given in the next subsection. It is clear that $\Lambda_e \subset \Lambda_b$. Thus, our proposed coding scheme instantiates the coset coding scheme introduced in [6], where the confidential message is mapped to the coset $\tilde{\lambda}_m \in \Lambda_b/\Lambda_e$.

⁷In this paper, a sequence of lattices Λ_e of increasing dimension is called secrecy-good if they achieve the strong secrecy capacity asymptotically. Note that this definition is different from that in [6], which is based on the flatness factor.

By using the above assignments and Lemma 3, we have

$$I(\mathbf{M}_\ell; \mathbf{Z}_\ell^{[N]}) \leq N2^{-N^{\beta'}}, \quad (14)$$

where $\mathbf{Z}_\ell^{[N]} = \mathbf{Z}^{[N]} \bmod \Lambda_\ell$. In other words, the employed polar code for the channel $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ can guarantee that the mutual information between the input message and the output is upper bounded by $N2^{-N^{\beta'}}$. According to Lemma 6, this polar code can also guarantee the same upper bound on the mutual information between the input message and the output of the channel $W'(\mathbf{X}_\ell; \mathbf{Z}|\mathbf{X}_{1:\ell-1})$ as shown in the following inequality (\mathbf{X}_ℓ is independent of the previous $\mathbf{X}_{1:\ell-1}$):

$$I(\mathbf{M}_\ell; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) \leq N2^{-N^{\beta'}}.$$

Recall $\mathbf{Z}^{[N]}$ is the signal received by Eve after the mod- Λ_r operation. From the chain rule of mutual information,

$$\begin{aligned} I(\mathbf{M}; \mathbf{Z}^{[N]}) &= \sum_{\ell=1}^r I(\mathbf{Z}^{[N]}; \mathbf{M}_\ell | \mathbf{M}_{1:\ell-1}) \\ &= \sum_{\ell=1}^r H(\mathbf{M}_\ell | \mathbf{M}_{1:\ell-1}) - H(\mathbf{M}_\ell | \mathbf{Z}^{[N]}, \mathbf{M}_{1:\ell-1}) \\ &\leq \sum_{\ell=1}^r H(\mathbf{M}_\ell) - H(\mathbf{M}_\ell | \mathbf{Z}^{[N]}, \mathbf{M}_{1:\ell-1}) \\ &= \sum_{\ell=1}^r I(\mathbf{M}_\ell; \mathbf{Z}^{[N]}, \mathbf{M}_{1:\ell-1}) \\ &\leq \sum_{\ell=1}^r I(\mathbf{M}_\ell; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) \leq rN2^{-N^{\beta'}}, \end{aligned} \quad (15)$$

where the last inequality holds because $I(\mathbf{M}_\ell; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) = I(\mathbf{M}_\ell; \mathbf{Z}^{[N]}, \mathbf{U}_{1:\ell-1}^{[N]})$ and adding more variables will not decrease the mutual information. Therefore strong secrecy is achieved since $\lim_{N \rightarrow \infty} I(\mathbf{M}; \mathbf{Z}^{[N]}) = 0$.

Remark 2: Note that the above analysis actually implies *semantic security*, i.e., (15) holds for arbitrarily distributed \mathbf{M} . This is because of the symmetric nature of the Λ_b/Λ_e channel [27]. Since the message \mathbf{M} is drawn from $\mathcal{R}(\Lambda_e)$ and the random bits are drawn from $\Lambda_e \cap \mathcal{R}(\Lambda_s)$, by Lemma 5, the mod- Λ_e map is information lossless and its output is a sufficient statistic for \mathbf{M} . In this sense, the channel between the confidential message and the Eavesdropper's signal can be viewed as a Λ_b/Λ_e channel. Since the Λ_b/Λ_e channel is symmetric, the maximum mutual information is achieved by the uniform input. Consequently, the mutual information corresponding to other input distributions can also be upper bounded by $rN2^{-N^{\beta'}}$ in (15). It is worth mentioning this Λ_b/Λ_e channel can be seen as the counterpart in lattice coding of the randomness-induced channel defined in [8].

Theorem 2 (Achieving secrecy capacity of the mod- Λ_s GWC): Consider a polar lattice L constructed according to (13) with the binary lattice partition chain $\Lambda/\Lambda_1/\dots/\Lambda_r$ and r binary nested polar codes with block length N . Scale Λ and r to satisfy the following conditions:

- (i) $h(\Lambda, \sigma_b^2) \rightarrow \log(\text{Vol}(\Lambda))$,

(ii) $h(\Lambda_r, \sigma_e^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_e^2)$.

Given $\sigma_e^2 > \sigma_b^2$, all strong secrecy rates R satisfying

$$R < \frac{1}{2} \log \frac{\sigma_e^2}{\sigma_b^2}$$

are achievable as $N \rightarrow \infty$, using the polar lattice L on the mod- Λ_s Gaussian wiretap channel.

Proof: By Lemma 4 and (13),

$$\begin{aligned} \lim_{N \rightarrow \infty} R &= \sum_{\ell=1}^r \lim_{N \rightarrow \infty} \frac{|\mathcal{A}_\ell|}{N} \\ &= \sum_{\ell=1}^r C(V_\ell) - C(W_\ell) \\ &= \sum_{\ell=1}^r C(V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)) - C(W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)) \\ &= C(V(\Lambda/\Lambda_r, \sigma_b^2)) - C(W(\Lambda/\Lambda_r, \sigma_e^2)) \\ &= C(\Lambda_r, \sigma_b^2) - C(\Lambda, \sigma_b^2) - C(\Lambda_r, \sigma_e^2) + C(\Lambda, \sigma_e^2) \\ &= h(\Lambda_r, \sigma_e^2) - h(\Lambda_r, \sigma_b^2) + h(\Lambda, \sigma_e^2) - h(\Lambda, \sigma_b^2) \\ &= \frac{1}{2} \log \frac{\sigma_e^2}{\sigma_b^2} - (\epsilon_e - \epsilon_b) - \epsilon_1, \end{aligned} \tag{16}$$

where

$$\begin{cases} \epsilon_1 = h(\Lambda, \sigma_e^2) - h(\Lambda, \sigma_b^2) \geq 0, \\ \epsilon_b = h(\sigma_b^2) - h(\Lambda_r, \sigma_b^2) = \frac{1}{2} \log(2\pi e \sigma_b^2) - h(\Lambda_r, \sigma_b^2) \geq 0, \\ \epsilon_e = h(\sigma_e^2) - h(\Lambda_r, \sigma_e^2) = \frac{1}{2} \log(2\pi e \sigma_e^2) - h(\Lambda_r, \sigma_e^2) \geq 0 \end{cases}$$

and $\epsilon_e - \epsilon_b \geq 0$.

By scaling Λ so that the mod- Λ noise is almost uniform, we can have $h(\Lambda, \sigma_b^2) \rightarrow \log(V(\Lambda))$. Since $\sigma_e^2 > \sigma_b^2$, we also have $h(\Lambda, \sigma_e^2) \rightarrow \log(V(\Lambda))$ and thus $\epsilon_1 \approx 0$. The number of levels is also increased until $h(\Lambda_r, \sigma_e^2) \approx \frac{1}{2} \log(2\pi e \sigma_e^2)$, hence $h(\Lambda_r, \sigma_b^2) \approx \frac{1}{2} \log(2\pi e \sigma_e^2)$, such that both ϵ_b and ϵ_e are almost 0. Therefore by scaling Λ_1 and adjusting r , the secrecy rate can get arbitrarily close to $\frac{1}{2} \log \frac{\sigma_e^2}{\sigma_b^2}$. \square

Remark 3: The secrecy capacity of the mod- Λ_s Gaussian wiretap channel per use is given by

$$C_s = \frac{1}{N} C(\Lambda_s, \sigma_b^2) - \frac{1}{N} C(\Lambda_s, \sigma_e^2) = \frac{1}{N} h(\Lambda_s, \sigma_e^2) - \frac{1}{N} h(\Lambda_s, \sigma_b^2)$$

since the wiretapper's channel is degraded with respect to the main channel. Because $h(\Lambda_r, \sigma_e^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_e^2)$ and $\Lambda_s \subset \Lambda_r^N$, we have $\frac{1}{N} h(\Lambda_s, \sigma_e^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_e^2)$ and $\frac{1}{N} h(\Lambda_s, \sigma_b^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_b^2)$. Hence $C_s \rightarrow \frac{1}{2} \log \frac{\sigma_e^2}{\sigma_b^2}$. It also equals the secrecy capacity of the Gaussian wiretap channel when the signal power goes to infinity. It is noteworthy that we successfully remove the $\frac{1}{2}$ -nat gap in the achievable secrecy rate derived in [6] which is caused by the limitation of the L^∞ distance associated with the flatness factor.

Remark 4: The mild conditions (i) and (ii) stated in the theorem are easy to meet, by scaling top lattice Λ and choosing the number of levels r appropriately. Consider an example for $\sigma_e^2 = 4$ and $\sigma_b^2 = 1$. We choose $r = 3$

levels and a partition chain $\mathbb{Z}/2\mathbb{Z}/4\mathbb{Z}$ with scaling factor 2.5. The difference between the achievable rate computed from (16) and the upper bound $\frac{1}{2} \log \frac{\sigma_e^2}{\sigma_b^2}$ on secrecy capacity is only 0.05.

Remark 5: From conditions (i) and (ii), we can see that the construction for secrecy-good lattices requires more levels than the construction of AWGN-good lattices. ϵ_1 can be made arbitrarily small by scaling down Λ such that both $h(\Lambda, \sigma_e^2)$ and $h(\Lambda, \sigma_b^2)$ are sufficiently close to $\log V(\Lambda)$. For polar lattices for AWGN-goodness [14], we only need $h(\Lambda_{r'}, \sigma_b^2) \approx \frac{1}{2} \log(2\pi e \sigma_b^2)$ for some $r' < r$. Since $\epsilon_b < \epsilon_e$, $\Lambda_{r'}$ may be not enough for the wiretapper's channel. Therefore, more levels are needed in the wiretap coding context. To satisfy the condition $h(\Lambda_r, \sigma_e^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_e^2)$, it is sufficient to guarantee that $P_e(\Lambda_r, \sigma_e^2) \rightarrow 0$ by [27, Theorem 13]. When one-dimensional binary partition $\mathbb{Z}/2\mathbb{Z}/4\mathbb{Z}/\dots$ is used, we have $P_e(\Lambda_r, \sigma_e^2) \leq Q(\frac{2^r}{2\sigma_e}) \leq e^{-\frac{2^{2r}}{8\sigma_e^2}}$, where $Q(\cdot)$ is the Q-function. Letting $r = O(\log N)$, the error probability vanishes as $P_e(\Lambda_r, \sigma_e^2) = e^{-O(N)}$, which implies that $h(\Lambda_r, \sigma_e^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_e^2)$ as $N \rightarrow \infty$.

D. Reliability

In the original polar coding scheme for the binary wiretap channel [8], how to assign set \mathcal{D} is a problem. Assigning frozen bits to \mathcal{D} guarantees reliability but only achieves weak secrecy, whereas assigning random bits to \mathcal{D} guarantees strong secrecy but may violate the reliability requirement because \mathcal{D} may be nonempty. In order to ensure strong secrecy, \mathcal{D} is assigned with random bits ($\mathcal{D} \leftarrow R$), which makes this scheme failed to accomplish the theoretical reliability. For any ℓ -th level channel $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ at Bob's end, the probability of error is upper bounded by the sum of the Bhattacharyya parameters $\tilde{Z}(V_N^{(j)}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2))$ of subchannels that are not frozen to zero. For each bit-channel index j and $\beta < 0.5$, we have

$$j \in \mathcal{G}(V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)) \cup \mathcal{D}_\ell.$$

By the definition (4), the sum of $\tilde{Z}(V_N^{(j)}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2))$ over the set $\mathcal{G}(V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2))$ is bounded by 2^{-N^β} , therefore the error probability of the ℓ -th level channel under the SC decoding, denoted by $P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$, can be upper bounded by

$$P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2) \leq N2^{-N^\beta} + \sum_{j \in \mathcal{D}_\ell} \tilde{Z}(V_N^{(j)}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)).$$

Since multistage decoding is utilized, by the union bound, the final decoding error probability for Bob is bounded as

$$\Pr\{\hat{M} \neq M\} \leq \sum_{i=1}^r P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2).$$

Unfortunately, a proof that this scheme satisfies the reliability condition cannot be attained here because the bound of the sum $\sum_{j \in \mathcal{D}_\ell} \tilde{Z}(V_N^{(j)}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2))$ is not known. Note that significantly low probabilities of error can still be achieved in practice since the size of \mathcal{D}_ℓ is very small for sufficiently large N .

The reliability problem was recently solved in [9], where a new scheme dividing the information message into several blocks was proposed. For a specific block, \mathcal{D}_ℓ is still assigned with random bits and transmitted in advance

in the set \mathcal{A}_ℓ of the previous block. This scheme involves negligible rate loss and finally realizes reliability and strong security simultaneously. In this case, if the reliability of each partition channel can be achieved, i.e., for any ℓ -th level partition $\Lambda_{\ell-1}/\Lambda_\ell$, $P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ vanishes as $N \rightarrow \infty$, then the total decoding error probability for Bob can be made arbitrarily small. Consequently, based on this new scheme of assigning the problematic set, the error probability on level ℓ can be upper bounded by

$$P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2) \leq \epsilon_{N'}^\ell + k_\ell \cdot O(2^{-N'^\beta}), \quad (17)$$

where k_ℓ is the number of information blocks on the ℓ -th level, N' is the length of each block which satisfies $N' \times k_\ell = N$ and $\epsilon_{N'}^\ell$ is caused by the first separate block on the ℓ -th level consisting of the initial bits in \mathcal{D}_ℓ . Since $|\mathcal{D}_\ell|$ is extremely small comparing to the block length N , the decoding failure probability for the first block can be made arbitrarily small when N is sufficiently large. Meanwhile, by the analysis in [15], when $h(\Lambda, \sigma_b^2) \rightarrow \log(V(\Lambda))$, $h(\Lambda_r, \sigma_b^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_b^2)$, and $R_C \rightarrow C(\Lambda/\Lambda_r, \sigma_b^2)$, we have $\gamma_{\Lambda_b}(\sigma_b) \rightarrow 2\pi e$. Therefore, Λ_b is an AWGN-good lattice⁸.

Note that the rate loss incurred by repeatedly transmitted bits in \mathcal{D}_ℓ is negligible because of its small size. Specifically, the actual secrecy rate in the ℓ -th level is given by $\frac{k_\ell}{k_\ell+1}[C(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2) - C(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)]$. Clearly, this rate can be made close to the secrecy capacity by choosing sufficiently large k_ℓ as well.

IV. SECRECY-GOOD POLAR LATTICES WITH DISCRETE GAUSSIAN SHAPING

In this section, we apply Gaussian shaping on the AWGN-good and secrecy-good polar lattices. The idea of lattice Gaussian shaping was proposed in [30] and then implemented in [15] to construct capacity-achieving polar lattices. For wiretap coding, the discrete Gaussian distribution can also be utilized to satisfy the power constraint. In simple terms, after obtaining the AWGN-good lattice Λ_b and the secrecy-good lattice Λ_e , Alice still maps each message m to a coset $\tilde{\lambda}_m \in \Lambda_b/\Lambda_e$ as mentioned in Sect. III. However, instead of the mod- Λ_s operation, Alice samples the encoded signal \mathbf{X}^N from $D_{\Lambda_e+\lambda_m, \sigma_s^2}$, where λ_m is the coset representative of $\tilde{\lambda}_m$ and σ_s^2 is arbitrarily close to the signal power P_s (see [6] for more details). The construction of polar lattices with Gaussian shaping is reviewed in Sect. IV-A. With Gaussian shaping, we propose a new partition of the index set for the genuine GWC in Sect. IV-B. Strong secrecy is proved in Sect. IV-C, and extension to semantical security is given in Sect. IV-D. Reliability is discussed in Sect. IV-E. Moreover, we will show that this shaping operation does not hurt the secrecy rate and that the strong secrecy capacity can be achieved.

A. Gaussian shaping over polar lattices

As shown in [15], the shaping scheme is based on the technique of polar codes for asymmetric channels. For the paper to be self-contained, a brief review will be presented in this subsection. A more detailed explanation of this Gaussian shaping technique can be found in [15].

⁸More precisely, to make Λ_b AWGN-good, we need $P_e(\Lambda_b, \sigma_b^2) \rightarrow 0$ by definition. By [15, Theorem 2], $P_e(\Lambda_b, \sigma_b^2) \leq rN2^{-N^\beta} + N \cdot P_e(\Lambda_r, \sigma_b^2)$. According to the analysis in Remark 5, $r = O(\log N)$ is sufficient to guarantee $P_e(\Lambda_r, \sigma_b^2) = e^{-O(N)}$, meaning that a sub-exponentially vanishing $P_e(\Lambda_b, \sigma_b^2)$ can be achieved.

Similarly to the polar coding on symmetric channels, the Bhattacharyya parameter for a binary memoryless asymmetric (BMA) channel is defined as follows.

Definition 3 (Bhattacharyya parameter for BMA channel): Let W be a BMA channel with input $X \in \mathcal{X} = \{0, 1\}$ and output $Y \in \mathcal{Y}$. The input distribution and channel transition probability is denoted by P_X and $P_{Y|X}$ respectively. The Bhattacharyya parameter Z for W is the defined as

$$\begin{aligned} Z(X|Y) &= 2 \sum_y P_Y(y) \sqrt{P_{X|Y}(0|y) P_{X|Y}(1|y)} \\ &= 2 \sum_y \sqrt{P_{X,Y}(0, y) P_{X,Y}(1, y)}. \end{aligned}$$

The following lemma, which will be useful for the forthcoming new partition scheme, shows that by adding observable at the output of W , Z will not decrease.

Lemma 7 (Conditioning reduces Bhattacharyya parameter Z [15]): Let $(X, Y, Y') \sim P_{X,Y,Y'}$, $X \in \mathcal{X} = \{0, 1\}$, $Y \in \mathcal{Y}$, $Y' \in \mathcal{Y}'$, we have

$$Z(X|Y, Y') \leq Z(X|Y).$$

When X is uniformly distributed, the Bhattacharyya parameter of BMA channels coincides with that of BMS channels defined in Definition 2. Moreover, the calculation of Z can be converted to the calculation of the Bhattacharyya parameter \tilde{Z} for a related BMS channel. The following lemma is implicitly considered in [36] and then explicitly expressed in [15]. We show it here for completeness.

Lemma 8 (From Asymmetric to Symmetric channel [15]): Let \tilde{W} be a binary input channel corresponding to the asymmetric channel W with input $\tilde{X} \in \mathcal{X} = \{0, 1\}$ and output $\tilde{Y} \in \{\mathcal{Y}, \mathcal{X}\}$. The input of \tilde{W} is uniformly distributed, i.e., $P_{\tilde{X}}(\tilde{x} = 0) = P_{\tilde{X}}(\tilde{x} = 1) = \frac{1}{2}$. The relationship between \tilde{W} and W is shown in Fig. 4. Then \tilde{W} is a binary symmetric channel in the sense that $P_{\tilde{Y}|\tilde{X}}(y, x \oplus \tilde{x} | \tilde{x}) = P_{Y,X}(y, x)$.

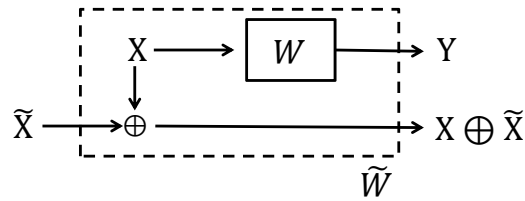


Fig. 4. The relationship between \tilde{W} and W .

The following lemma describes how to construct a polar code for a BMA channel W from that for the associated BMS channel \tilde{W} .

Lemma 9 (The equivalence between symmetric and asymmetric Bhattacharyya parameters [36]): For a BMA channel W with input $X \sim P_X$, let \tilde{W} be its symmetrized channel constructed according to Lemma 8. Suppose $X^{[N]}$ and $Y^{[N]}$ be the input and output vectors of W^N , and let $\tilde{X}^{[N]}$ and $\tilde{Y}^{[N]} = (X^{[N]} \oplus \tilde{X}^{[N]}, Y^{[N]})$ be the input

and output vectors of \tilde{W}^N . Consider polarized random variables $U^{[N]} = X^{[N]}G_N$ and $\tilde{U}^{[N]} = \tilde{X}^{[N]}G_N$, and denote by W_N and \tilde{W}_N the combining channel of N uses of W and \tilde{W} , respectively. The Bhattacharyya parameter for each subchannel of W_N is equal to that of each subchannel of \tilde{W}_N , i.e.,

$$Z(U^i|U^{1:i-1}, Y^{[N]}) = \tilde{Z}(\tilde{U}^i|\tilde{U}^{1:i-1}, X^{[N]} \oplus \tilde{X}^{[N]}, Y^{[N]}).$$

To obtain the desired input distribution of P_X for W , the indices with very small $Z(U^i|U^{1:i-1})$ should be removed from the information set of the symmetric channel. Following [15], the resultant subset is referred to as the information set \mathcal{I} for the asymmetric channel W . For the remaining part \mathcal{I}^c , we further find out that there are some bits which can be made independent of the information bits and uniformly distributed. The purpose of extracting such bits is for the interest of our lattice construction. We name the set that includes those independent frozen bits as the frozen set \mathcal{F} , and the remaining bits are determined by the bits in $\mathcal{F} \cup \mathcal{I}$. We name the set of all those deterministic bits as the shaping set \mathcal{S} . The three sets are formally defined as follows:

$$\begin{cases} \text{frozen set: } \mathcal{F} = \{i \in [N] : Z(U^i|U^{1:i-1}, Y^{[N]}) \geq 1 - 2^{-N^\beta}\} \\ \text{information set: } \mathcal{I} = \{i \in [N] : Z(U^i|U^{1:i-1}, Y^{[N]}) \leq 2^{-N^\beta} \text{ and } Z(U^i|U^{1:i-1}) \geq 1 - 2^{-N^\beta}\} \\ \text{shaping set: } \mathcal{S} = (\mathcal{F} \cup \mathcal{I})^c. \end{cases} \quad (18)$$

To identify these three sets, one can use Lemma 9 to calculate $Z(U^i|U^{1:i-1}, Y^{[N]}, X^{[N]})$ using the known constructing techniques for symmetric polar codes [33] [37]. We note that $Z(U^i|U^{1:i-1})$ can be computed in a similar way, by constructing a symmetric channel between \tilde{X} and $X \oplus \tilde{X}$. Besides the construction, the decoding process for the asymmetric polar codes can also be converted to the decoding for the symmetric polar codes.

The polar coding scheme according to (18), which can be viewed as an extension of the scheme proposed in [36], has been proved to be capacity-achieving in [15]. Moreover, it can be extended to the construction of multilevel asymmetric polar codes.

Theorem 3 (Construction of multilevel polar codes [15]): Consider a polar code with the following encoding strategy for the channel of the ℓ -th ($\ell \leq r$) level W_ℓ with the channel transition probability $P_{Y|X_\ell, X_{1:\ell-1}}(y|x_\ell, x_{1:\ell-1})$.

- Encoding: Before sending the codeword $x_\ell^{[N]} = u_\ell^{[N]}G_N$, the index set $[N]$ are divided into three parts: the frozen set \mathcal{F}_ℓ , information set \mathcal{I}_ℓ , and shaping set \mathcal{S}_ℓ , which are defined as follows:

$$\begin{cases} \mathcal{F}_\ell = \{i \in [N] : Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}, Y^{[N]}) \geq 1 - 2^{-N^\beta}\} \\ \mathcal{I}_\ell = \{i \in [N] : Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}, Y^{[N]}) \leq 2^{-N^\beta} \text{ and } Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) \geq 1 - 2^{-N^\beta}\} \\ \mathcal{S}_\ell = (\mathcal{F}_\ell \cup \mathcal{I}_\ell)^c. \end{cases}$$

The encoder first places uniformly distributed information bits in \mathcal{I}_ℓ . Then the frozen set \mathcal{F}_ℓ is filled with a uniform random sequence which is shared between the encoder and the decoder. The bits in \mathcal{S}_ℓ are generated by a random mapping $\Phi_{\mathcal{S}_\ell}$, which yields the following distribution:

$$u_\ell^i = \begin{cases} 0 & \text{with probability } P_{U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}}(0|u_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}), \\ 1 & \text{with probability } P_{U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}}(1|u_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}). \end{cases} \quad (19)$$

Then any message rate arbitrarily close to $I(\mathbf{X}_\ell; \mathbf{Y} | \mathbf{X}_{1:\ell-1})$ is achievable using the SC decoding⁹ and the expectation of the decoding error probability over the randomized mappings satisfies $E_{\Phi_{S_\ell}}[P_e(\phi_{S_\ell})] = O(2^{-N^{\beta'}})$ for any $\beta' < \beta < 0.5$.

Now let us pick a suitable input distribution $P_{\mathbf{X}_{1:r}}$ to implement the shaping. As shown in Theorem 1, the mutual information between the discrete Gaussian lattice distribution D_{Λ, σ_s} and the output of the AWGN channel approaches $\frac{1}{2} \log(1 + \text{SNR})$ as the flatness factor $\epsilon_\Lambda(\tilde{\sigma}) \rightarrow 0$. Therefore, we use the lattice Gaussian distribution $P_{\mathbf{X}} \sim D_{\Lambda, \sigma_s}$ as the constellation, which gives us $\lim_{r \rightarrow \infty} P_{\mathbf{X}_{1:r}} = P_{\mathbf{X}} \sim D_{\Lambda, \sigma_s}$. By [15, Lemma 5], when $N \rightarrow \infty$, the mutual information $I(\mathbf{X}_r; \mathbf{Y} | \mathbf{X}_{1:r-1})$ at the bottom level goes to 0 if $r = O(\log \log N)$, and using the first r levels would involve a capacity loss $\sum_{\ell > r} I(\mathbf{X}_\ell; \mathbf{Y} | \mathbf{X}_{1:\ell-1}) \leq O(\frac{1}{N})$.

From the chain rule of mutual information,

$$I(\mathbf{X}_{1:r}; \mathbf{Y}) = \sum_{\ell=1}^r I(\mathbf{X}_\ell; \mathbf{Y} | \mathbf{X}_{1:\ell-1}),$$

we have r binary-input channels and the ℓ -th channel according to $I(\mathbf{X}_\ell; \mathbf{Y} | \mathbf{X}_{1:\ell-1})$ is generally asymmetric with the input distribution $P_{\mathbf{X}_\ell | \mathbf{X}_{1:\ell-1}}$ ($1 \leq \ell \leq r$). Then we can construct the polar code for the asymmetric channel at each level according to Lemma 8. It is shown in [15] that the ℓ -th symmetrized channel is equivalent to the MMSE-scaled $\Lambda_{\ell-1}/\Lambda_\ell$ channel in the sense of channel polarization.

Therefore, when power constrain is taken into consideration, the multilevel polar codes before shaping are constructed according to the symmetric channel $V(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}_b^2)$ and $W(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}_e^2)$, where $\tilde{\sigma}_b^2 = \left(\frac{\sigma_s \sigma_b}{\sqrt{\sigma_s^2 + \sigma_b^2}}\right)^2$ and $\tilde{\sigma}_e^2 = \left(\frac{\sigma_s \sigma_e}{\sqrt{\sigma_s^2 + \sigma_e^2}}\right)^2$ are the MMSE-scaled noise variance of the main channel and of the wiretapper's channel, respectively. This is similar to the mod- Λ_s GWC scenario mentioned in the previous section. The difference is that σ_b^2 and σ_e^2 are replaced by $\tilde{\sigma}_b^2$ and $\tilde{\sigma}_e^2$ accordingly. As a result, we can still obtain an AWGN-good lattice Λ_b and a secrecy-good lattice Λ_e by treating $V(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}_b^2)$ and $W(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}_e^2)$ as the main channel and wiretapper's channel at each level.

B. Three-dimensional partition

Now we consider the partition of the index set $[N]$ with shaping involved. According to the analysis of asymmetric polar codes, we have to eliminate those indices with small $Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]})$ from the information set of the symmetric channels. Therefore, Alice cannot send message on those subchannels with $Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta}$. Note that this part is the same for \tilde{V}_ℓ and \tilde{W}_ℓ , because it only depends on the shaping distribution. At each level, the index set which is used for shaping is given as

$$\mathcal{S}_\ell \triangleq \{i \in [N] : Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta}\},$$

⁹It is possible to derandomize the mapping Φ_{S_ℓ} for the purpose of achieving capacity alone. However, it is tricky to handle the random mapping in order to achieve the secrecy capacity: it requires either to share a secret random mapping or to use the Markov block coding technique (see Sect. IV-E).

and the index set which is not for shaping is denoted by \mathcal{S}_ℓ^c . Recall that for the index set $[N]$, we already have two partition criteria, i.e., reliability-good and information-bad (see (4)). We rewrite the reliability-good index set \mathcal{G}_ℓ and information-bad index set \mathcal{N}_ℓ at level ℓ as

$$\begin{aligned}\mathcal{G}_\ell &\triangleq \{i \in [N] : Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{Y}^{[N]}) \leq 2^{-N^\beta}\}, \\ \mathcal{N}_\ell &\triangleq \{i \in [N] : Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{Z}^{[N]}) \geq 1 - 2^{-N^\beta}\}.\end{aligned}\quad (20)$$

Note that \mathcal{G}_ℓ and \mathcal{N}_ℓ are defined by the asymmetric Bhattacharyya parameters. Nevertheless, by Lemma 9 and the channel equivalence, we have $\mathcal{G}_\ell = \mathcal{G}(\tilde{V}_\ell)$ and $\mathcal{N}_\ell = \mathcal{N}(\tilde{W}_\ell)$ as defined in (4), where \tilde{V}_ℓ and \tilde{W}_ℓ are the respective symmetric channels or the MMSE-scaled $\Lambda_{\ell-1}/\Lambda_\ell$ channels for Bob and Eve at level ℓ . The four sets \mathcal{A}_ℓ , \mathcal{B}_ℓ , \mathcal{C}_ℓ , and \mathcal{D}_ℓ are defined in the same fashion as (5), with \mathcal{G}_ℓ and \mathcal{N}_ℓ replacing $\mathcal{G}(\tilde{V}_\ell)$ and $\mathcal{N}(\tilde{W}_\ell)$, respectively. Now the whole index set $[N]$ is divided like a cube in three directions, which is shown in Fig. 5.

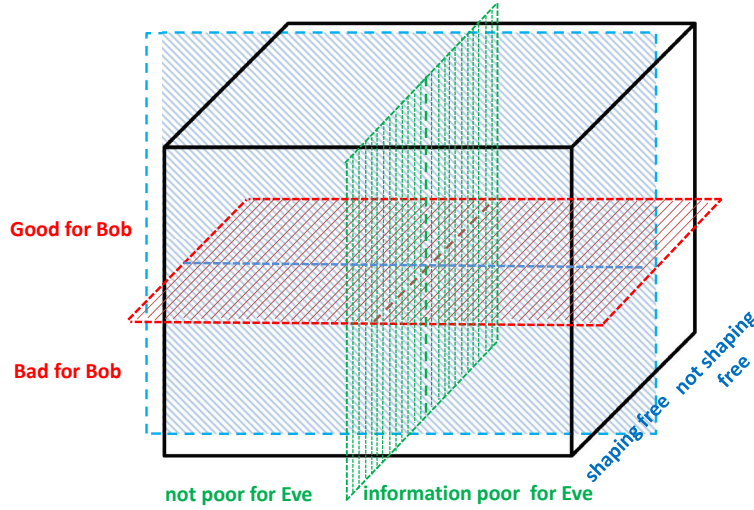


Fig. 5. Partitions of the index set $[N]$ with shaping.

Clearly, we have eight blocks:

$$\begin{aligned}\mathcal{A}_\ell^S &= \mathcal{A}_\ell \cap \mathcal{S}_\ell, \quad \mathcal{A}_\ell^{S^c} = \mathcal{A}_\ell \cap \mathcal{S}_\ell^c \\ \mathcal{B}_\ell^S &= \mathcal{B}_\ell \cap \mathcal{S}_\ell, \quad \mathcal{B}_\ell^{S^c} = \mathcal{B}_\ell \cap \mathcal{S}_\ell^c \\ \mathcal{C}_\ell^S &= \mathcal{C}_\ell \cap \mathcal{S}_\ell, \quad \mathcal{C}_\ell^{S^c} = \mathcal{C}_\ell \cap \mathcal{S}_\ell^c \\ \mathcal{D}_\ell^S &= \mathcal{D}_\ell \cap \mathcal{S}_\ell, \quad \mathcal{D}_\ell^{S^c} = \mathcal{D}_\ell \cap \mathcal{S}_\ell^c\end{aligned}\quad (21)$$

By Lemma 7, we observe that $\mathcal{A}_\ell^S = \mathcal{C}_\ell^S = \emptyset$, $\mathcal{A}_\ell^{S^c} = \mathcal{A}_\ell$, and $\mathcal{C}_\ell^{S^c} = \mathcal{C}_\ell$. The shaping set \mathcal{S}_ℓ is divided into two sets \mathcal{B}_ℓ^S and \mathcal{D}_ℓ^S . The bits in \mathcal{S}_ℓ are determined by the bits in \mathcal{S}_ℓ^c according to the mapping. Similarly, \mathcal{S}_ℓ^c is divided into the four sets $\mathcal{A}_\ell^{S^c} = \mathcal{A}_\ell$, $\mathcal{B}_\ell^{S^c}$, $\mathcal{C}_\ell^{S^c} = \mathcal{C}_\ell$, and $\mathcal{D}_\ell^{S^c}$. Note that for wiretap coding, the frozen set becomes $\mathcal{C}_\ell^{S^c}$, which is slightly different from the frozen set for channel coding. To satisfy the reliability condition, the frozen set $\mathcal{C}_\ell^{S^c}$ and the problematic set $\mathcal{D}_\ell^{S^c}$ cannot be set uniformly random any more. Recall that only the independent frozen

set \mathcal{F}_ℓ at each level, which is defined as $\{i \in [N] : Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{Y}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) \geq 1 - 2^{-N^\beta}\}$, can be set uniformly random (which are already shared between Alice and Bob), and the bits in the unpolarized frozen set $\bar{\mathcal{F}}_\ell$, defined as $\{i \in [N] : 2^{-N^\beta} < Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{Y}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta}\}$, should be determined according to the mapping. Moreover, we can observe that $\mathcal{F}_\ell \subset \mathcal{C}_\ell^{S^c}$ and $\mathcal{D}_\ell^{S^c} \subset \mathcal{D}_\ell \subset \bar{\mathcal{F}}_\ell$. Here we make the bits in \mathcal{F}_ℓ uniformly random and the bits in $\mathcal{C}_\ell^{S^c} \setminus \mathcal{F}_\ell$ and $\mathcal{D}_\ell^{S^c}$ determined by the mapping. Therefore, from now on, we adjust the definition of the shaping bits as:

$$\mathcal{S}_\ell \triangleq \{i \in [N] : Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta} \text{ or } 2^{-N^\beta} < Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{Y}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta}\}, \quad (22)$$

which is essentially equivalent to the definition of the shaping set given in Theorem 3.

To sum up, at level ℓ , we assign the sets $\mathcal{A}_\ell^{S^c}$, $\mathcal{B}_\ell^{S^c}$, and \mathcal{F}_ℓ with message bits M_ℓ , uniformly random bits R_ℓ , and uniform frozen bits F_ℓ , respectively. The rest bits S_ℓ (in \mathcal{S}_ℓ) will be fed with random bits according to $P_{\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}}$. Clearly, this shaping operation will make the input distribution arbitrarily close to $P_{\mathbf{X}_\ell | \mathbf{X}_{1:\ell-1}}$. In this case, we can obtain the equality between the Bhattacharyya parameter of asymmetric setting and symmetric setting (see Lemma 9). This provides us a convenient way to prove the strong secrecy of the wiretap coding scheme with shaping because we have already proved the strong secrecy of a symmetric wiretap coding scheme using the Bhattacharyya parameter of the symmetric setting. A detailed proof will be presented in the following subsection. Before this, we show that the shaping will not change the message rate.

Lemma 10: For the symmetrized main channel \tilde{V}_ℓ and wiretapper's channel \tilde{W}_ℓ , consider the reliability-good indices set \mathcal{G}_ℓ and information-bad indices set \mathcal{N}_ℓ defined as in (20). By eliminating the shaping set \mathcal{S}_ℓ from the original message set defined in (5), we get the new message set $\mathcal{A}_\ell^{S^c} = \mathcal{G}_\ell \cap \mathcal{N}_\ell \cap \mathcal{S}_\ell^c$. The proportion of $|\mathcal{A}_\ell^{S^c}|$ equals to that of $|\mathcal{A}_\ell|$, and the message rate after shaping can still be arbitrarily close to $\frac{1}{2} \log \frac{\tilde{\sigma}_a^2}{\tilde{\sigma}_b^2}$.

Proof: By Theorem 2, when shaping is not involved, the message rate can be made arbitrarily close to $\frac{1}{2} \log \frac{\tilde{\sigma}_a^2}{\tilde{\sigma}_b^2}$. By the new definition (22) of \mathcal{S}_ℓ , we still have $\mathcal{A}_\ell^S = \emptyset$, which means the shaping operation will not affect the message rate. \square

C. Strong secrecy

In this subsection, we introduce a new induced channel from Eve's perspective and prove that the information leakage over this channel is vanishing at each level in Lemma 11. Then, strong secrecy is proved by using the chain rule of mutual information as in (15).

In [8], an induced channel is defined in order to prove strong secrecy. Here we call it the randomness-induced channel because it is induced by feeding the subchannels in the sets \mathcal{B}_ℓ and \mathcal{D}_ℓ with uniformly random bits. However, when shaping is involved, the set \mathcal{B}_ℓ and \mathcal{D}_ℓ are no longer fed with uniformly random bits. In fact, some subchannels (covered by the shaping mapping) should be fed with bits according to a random mapping. We define the channel induced by the shaping bits as the shaping-induced channel.

Definition 4 (Shaping-induced channel): The shaping-induced channel $\mathcal{Q}_N(W, \mathcal{S})$ is defined in terms of N uses of an asymmetric channel W , and a shaping subset \mathcal{S} of $[N]$ of size $|\mathcal{S}|$. The input alphabet of $\mathcal{Q}_N(W, \mathcal{S})$ is

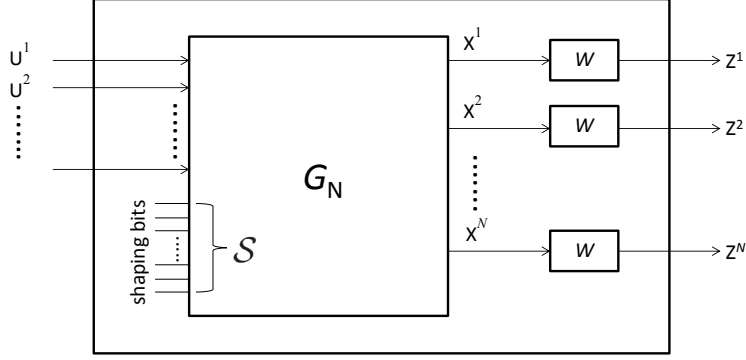


Fig. 6. Block diagram of the shaping-induced channel $\mathcal{Q}_N(W, S)$.

$\{0, 1\}^{N-|S|}$ and the bits in S are determined by the input bits according to a random shaping Φ_S . A block diagram of the shaping induced channel is shown in Fig. 6.

Based on the shaping-induced channel, we define a new induced channel, which is caused by feeding a part of the input bits of the shaping-induced channel with uniformly random bits.

Definition 5 (New induced channel): Based on a shaping induced channel $\mathcal{Q}_N(W, S)$, the new induced channel $\mathcal{Q}_N(W, S, \mathcal{R})$ is specified in terms of a randomness subset \mathcal{R} of size $|\mathcal{R}|$. The randomness is introduced into the input set of the shaping-induced channel. The input alphabet of $\mathcal{Q}_N(W, S, \mathcal{R})$ is $\{0, 1\}^{N-|S|-|\mathcal{R}|}$ and the bits in \mathcal{R} are uniformly and independently random. A block diagram of the new induced channel is shown in Fig. 7.

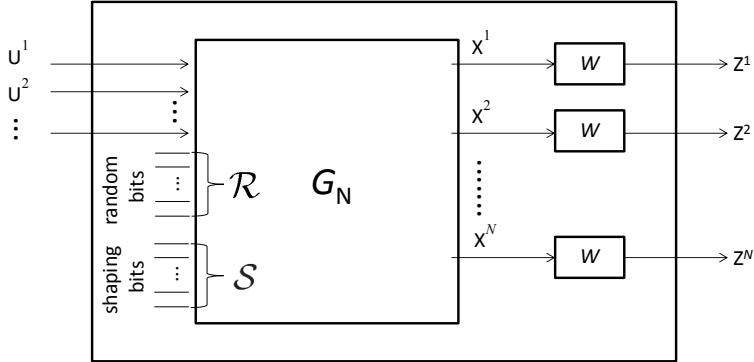


Fig. 7. Block diagram of the new induced channel $\mathcal{Q}_N(W, S, \mathcal{R})$.

The new induced channel is a combination of the shaping-induced channel and randomness-induced channel. This is different from the definition given in [8] because the bits in S are neither independent to the message bits nor uniformly distributed. As long as the input bits of the new induced channel are uniform and the shaping bits are chosen according to the random mapping, the new induced channel can still generate 2^N possible realizations $x_\ell^{[N]}$ of $\mathbf{X}_\ell^{[N]}$ as N goes to infinity, and those $x_\ell^{[N]}$ can be viewed as the output of N i.i.d binary sources

with input distribution $P_{X_\ell|X_{1:\ell-1}}$. These are exactly the conditions required by Lemma 9. Specifically, we have $Z(U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}, Z^{[N]}) = \tilde{Z}(\tilde{U}_\ell^i|\tilde{U}_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}, X_\ell^{[N]} \oplus \tilde{X}_\ell^{[N]}, Z^{[N]})$. In simple words, this equation holds when $x_\ell^{[N]}$ and $x_\ell^{[N]} \oplus \tilde{x}_\ell^{[N]}$ are all selected from $\{0, 1\}^N$ according to their respective distributions. Then we can exploit the relation between the asymmetric channel and the corresponding symmetric channel to bound the mutual information of the asymmetric channel. Therefore, we have to stick to the input distribution (uniform) of our new induced channel and also the distribution of the random mapping. This is similar to the setting of the randomness induced channel in [8], where the input distribution and the randomness distribution are both set to be uniform. In [8], the randomness-induced channel is further proved to be symmetric; then any other input distribution can also achieve strong secrecy and the symmetry finally results in semantic security. In this work, however, we do not have a proof of the symmetry of the new induced channel. For this reason, we assume for now that the message bits are uniform distributed. To prove semantic security, we will show that the information leakage of the symmetrized version of the new induced channel is vanishing in Sect. IV-D.

Lemma 11: Let M_ℓ be the uniformly distributed message bits and F_ℓ be the independent frozen bits at the input of the channel at the ℓ -th level. When shaping bits S_ℓ are selected according to the random mapping Φ_{S_ℓ} ¹⁰ and N is sufficiently large, the mutual information can be upper-bounded as

$$I(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) \leq O(N^2 2^{-N^{\beta'}}).$$

Proof: We firstly assume that U_ℓ^i is selected according to the distribution $P_{U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}}$ for all $i \in [N]$, i.e.,

$$u_\ell^i = \begin{cases} 0 & \text{with probability } P_{U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}}(0|u_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}), \\ 1 & \text{with probability } P_{U_\ell^i|U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}}(1|u_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}). \end{cases} \quad (23)$$

for all $i \in [N]$. In this case, the input distribution $P_{X_\ell|X_{1:\ell-1}}$ at each level is exactly the optimal input distribution obtained from the lattice Gaussian distribution. The mutual information between $M_\ell F_\ell$ and $(Z^{[N]}, X_{1:\ell-1}^{[N]})$ in this case is denoted by $I_P(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]})$.

For the shaping induced channel $\mathcal{Q}_N(W_\ell, S_\ell, \mathcal{R}_\ell)$ (\mathcal{R}_ℓ is $\mathcal{B}_\ell^{S_\ell^c}$ according to the above analysis), we write the indices of the input bits $(S_\ell \cup \mathcal{R}_\ell)^c = [N] \setminus (S_\ell \cup \mathcal{R}_\ell)$ as $\{i_1, i_2, \dots, i_{N-s_\ell-r_\ell}\}$, where $|\mathcal{R}_\ell| = r_\ell$ and $|S_\ell| = s_\ell$, and

¹⁰As we will see in Sect. IV-E, to achieve reliability, Alice needs to secretly share Φ_{S_ℓ} with Bob, or to use the Markov block coding technique.

assume that $i_1 < i_2 < \dots < i_{N-s_\ell-r_\ell}$. We have

$$\begin{aligned}
I_P(\mathbf{M}_\ell \mathbf{F}_\ell; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) &= I_P(\mathbf{U}_\ell^{(\mathcal{S}_\ell \cup \mathcal{R}_\ell)^c}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) \\
&= I_P(\mathbf{U}_\ell^{i_1}, \mathbf{U}_\ell^{i_2}, \dots, \mathbf{U}_\ell^{i_{N-r_\ell-s_\ell}}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) \\
&= \sum_{j=1}^{N-r_\ell-s_\ell} I_P(\mathbf{U}_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]} | \mathbf{U}_\ell^{i_1}, \mathbf{U}_\ell^{i_2}, \dots, \mathbf{U}_\ell^{i_{j-1}}) \\
&= \sum_{j=1}^{N-r_\ell-s_\ell} I_P(\mathbf{U}_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{U}_\ell^{i_1}, \mathbf{U}_\ell^{i_2}, \dots, \mathbf{U}_\ell^{i_{j-1}}) \\
&\stackrel{(a)}{\leq} \sum_{j=1}^{N-r_\ell-s_\ell} I_P(\mathbf{U}_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{U}_\ell^1, \mathbf{U}_\ell^2, \dots, \mathbf{U}_\ell^{i_{j-1}}),
\end{aligned}$$

where (a) holds because adding more variables will not decrease the mutual information.

Then the above mutual information can be bounded by the mutual information of the symmetric channel plus an infinitesimal term as follows:

$$\begin{aligned}
&\sum_{j=1}^{N-r_\ell-s_\ell} I_P(\mathbf{U}_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{U}_\ell^{1:i_j-1}) \\
&\stackrel{(a)}{\leq} \sum_{j=1}^{N-r_\ell-s_\ell} I(\tilde{\mathbf{U}}_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \tilde{\mathbf{X}}_\ell^{[N]} \oplus \mathbf{X}_\ell^{[N]}, \tilde{\mathbf{U}}_\ell^{1:i_j-1}) + H(\tilde{\mathbf{U}}_\ell^{i_j} | \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \tilde{\mathbf{X}}_\ell^{[N]} \oplus \mathbf{X}_\ell^{[N]}, \tilde{\mathbf{U}}_\ell^{1:i_j-1}) \\
&\quad - \sum_{j=1}^{N-r_\ell-s_\ell} H(\mathbf{U}_\ell^{i_j} | \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{U}_\ell^{1:i_j-1}) \\
&\stackrel{(b)}{\leq} \sum_{j=1}^{N-r_\ell-s_\ell} I(\tilde{\mathbf{U}}_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \tilde{\mathbf{X}}_\ell^{[N]} \oplus \mathbf{X}_\ell^{[N]}, \tilde{\mathbf{U}}_\ell^{1:i_j-1}) \\
&\quad + \sum_{j=1}^{N-r_\ell-s_\ell} Z(\mathbf{U}_\ell^{i_j} | \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{U}_\ell^{1:i_j-1}) - (Z(\mathbf{U}_\ell^{i_j} | \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{U}_\ell^{1:i_j-1}))^2 \\
&\stackrel{(c)}{\leq} \sum_{j=1}^{N-r_\ell-s_\ell} I(\tilde{\mathbf{U}}_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \tilde{\mathbf{X}}_\ell^{[N]} \oplus \mathbf{X}_\ell^{[N]}, \tilde{\mathbf{U}}_\ell^{1:i_j-1}) + N2^{-N^\beta} \\
&\stackrel{(d)}{\leq} N2^{-N^{\beta'}} + N2^{-N^\beta} \\
&\leq 2N2^{-N^{\beta'}}
\end{aligned}$$

for $0 < \beta' < \beta < 0.5$. Inequalities (a)-(d) follow from

- (a) uniformly distributed $\tilde{\mathbf{U}}_\ell^{i_j}$,
- (b) [38, Proposition 2] which gives $H(\mathbf{X}|\mathbf{Y}) - H(\mathbf{X}|\mathbf{Y}, \mathbf{Z}) \leq Z(\mathbf{X}|\mathbf{Y}) - (Z(\mathbf{X}|\mathbf{Y}, \mathbf{Z}))^2$ and Lemma 9,
- (c) our coding scheme guaranteeing that $Z(\mathbf{U}_\ell^{i_j} | \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{U}_\ell^{1:i_j-1})$ is greater than $1 - 2^{-N^\beta}$ for the frozen bits and information bits,
- (d) Lemma 2.

For wiretap coding, the message \mathbf{M}_ℓ , frozen bits \mathbf{F}_ℓ and random bits \mathbf{R}_ℓ are all uniformly random, and the shaping bits \mathbf{S}_ℓ are determined by \mathbf{S}_ℓ^c according to $\Phi_{\mathcal{S}_\ell}$. Let $Q_{\mathbf{U}_\ell^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{Z}^{[N]}}$ denote the joint distribution of

$(U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]})$ resulted from uniformly distributed $M_\ell F_\ell R_\ell$ and S_ℓ according to Φ_{S_ℓ} . By the proofs of [15, Th. 5] and [15, Th. 6], the total variation distance can be bounded as

$$\|Q_{U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]}} - P_{U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]}}\| \leq N2^{-N^{\beta'}} \quad (24)$$

for sufficiently large N .

By [39, Proposition 5], the mutual information $I(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]})$ caused by $Q_{U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]}}$ satisfies

$$\begin{aligned} \left| I(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) - I_P(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) \right| &\leq 7N2^{-N^{\beta'}} \log 2^N + h_2(N2^{-N^{\beta'}}) + h_2(4N2^{-N^{\beta'}}) \\ &= O(N2^{-N^{\beta'}}), \end{aligned}$$

where $h_2(\cdot)$ denotes the binary entropy function. □

Finally, strong secrecy (for uniform message bits) can be proved in the same fashion as shown in (15) as:

$$I(M; Z^{[N]}) \leq \sum_{\ell=1}^r I(M_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) \leq \sum_{\ell=1}^r I(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) = O(rN2^{-N^{\beta'}}).$$

Therefore we conclude that the whole shaping scheme is secure in the sense that the mutual information leakage between M and $Z^{[N]}$ vanishes with the block length N .

D. Semantic security

In this subsection, we extend strong secrecy of the constructed polar lattices to semantic security, namely the resulted strong secrecy does not rely on the distribution of the message. We take the level-1 wiretapper's channel W_1 as an example. Our goal is to show that the maximum mutual information between $M_1 F_1$ and $Z^{[N]}$ is vanishing for any input distribution as $N \rightarrow \infty$. Unlike the symmetric randomness induced channel introduced in [8], the new induced channel is generally asymmetric with transition probability

$$Q(z|v) = \frac{1}{2^{r_1}} \sum_{\Phi_{S_1}} P(\Phi_{S_1}) \sum_{e \in \{0,1\}^{r_1}} W_1^N(z|(v, e, \Phi_{S_1}(v, e))G_N),$$

where $\Phi_{S_1}(v, e)$ represents the shaping bits determined by v (the frozen bits and message bits together) and e (the random bits) according to the random mapping Φ_{S_1} . It is difficult to find the optimal input distribution to maximize the mutual information for the new induced channel.

To prove the semantic security, we investigate the relationship between the i -th subchannel of $W_{1,N}$ and the i -th subchannel of its symmetrized version $\tilde{W}_{1,N}$, which are denoted by $W_1^{(i,N)}$ and $\tilde{W}_1^{(i,N)}$, respectively. According to Lemma 8, the asymmetric wiretap channel $W_1 : X_1 \rightarrow Z$ is symmetrized to channel $\tilde{W}_1 : \tilde{X}_1 \rightarrow (Z, \tilde{X}_1 \oplus X_1)$. After the N -by- N polarization transform, we obtain $W_1^{(i,N)} : U_1^i \rightarrow (U_1^{1:i-1}, Z^{[N]})$ and $\tilde{W}_1^{(i,N)} : \tilde{U}_1^i \rightarrow (\tilde{U}_1^{1:i-1}, \tilde{X}_1^{[N]} \oplus X_1^{[N]}, Z^{[N]})$. The next lemma shows that if we symmetrize $W_1^{(i,N)}$ directly, i.e., construct a symmetric channel $\widetilde{W}_1^{(i,N)} : \tilde{U}_1^i \rightarrow (U_1^{1:i-1}, Z^{[N]}, \tilde{U}_1^i \oplus U_1^i)$ in the sense of Lemma 8, $W_1^{(i,N)}$ is degraded with respect to $\tilde{W}_1^{(i,N)}$.

Lemma 12: The symmetrized channel $\widetilde{W}_1^{(i,N)}$ derived directly from $W_1^{(i,N)}$ is degraded with respect to the i -th subchannel $\tilde{W}_1^{(i,N)}$ of \tilde{W}_1 .

Proof: According to the proof of [36, Theorem 2], we have the relationship

$$\tilde{W}_1^{(i,N)}(\tilde{u}_1^{1:i-1}, \tilde{x}_1^{[N]} \oplus x_1^{[N]}, z^{[N]} | \tilde{u}_1^i) = 2^{-N+1} P_{U_1^{1:i}, Z^{[N]}}(u_1^{1:i}, z^{[N]}).$$

Letting $\tilde{x}_1^{[N]} \oplus x_1^{[N]} = 0^{[N]}$, the equation becomes $\tilde{W}_1^{(i,N)}(u_1^{1:i-1}, 0^{[N]}, z^{[N]} | u_1^i) = 2^{-N+1} P_{U_1^{1:i}, Z^{[N]}}(u_1^{1:i}, z^{[N]})$, which has already been addressed in [36]. However, for a fixed $x_1^{[N]}$ and $\tilde{u}_1^i = u_1^i$, since G_N is full rank, there are 2^{N-1} choices of $\tilde{x}_1^{[N]}$ remaining, which means that there exists 2^{N-1} outputs symbols of $\tilde{W}_1^{(i,N)}$ having the same transition probability $2^{-N+1} P_{U_1^{1:i}, Z^{[N]}}(u_1^{1:i}, z^{[N]})$. Suppose a middle channel which maps all these output symbols to one single symbol, which is with transition probability $P_{U_1^{1:i}, Z^{[N]}}(u_1^{1:i}, z^{[N]})$. The same operation can be done for $\tilde{u}_1^i = u_1^i \oplus 1$, making another symbol with transition probability $P_{U_1^{1:i}, Z^{[N]}}(u_1^{1:i}, z^{[N]})$ corresponding to the input $u_1^i \oplus 1$. This is a channel degradation process, and the degraded channel is symmetric.

Then we show that the symmetrized channel $\widetilde{W_1^{(i,N)}}$ is equivalent to the degraded channel mentioned above. By Lemma 8, the channel transition probability of $\widetilde{W_1^{(i,N)}}$ is

$$\widetilde{W_1^{(i,N)}}(u_1^{1:i-1}, \tilde{u}_1^i \oplus u_1^i, z^{[N]} | \tilde{u}_1^i) = P_{U_1^{1:i}, Z^{[N]}}(u_1^{1:i}, z^{[N]}),$$

which is equal to the transition probability of the degraded channel discussed in the previous paragraph. Therefore, $\widetilde{W_1^{(i,N)}}$ is degraded with respect to $\tilde{W}_1^{(i,N)}$. \square

Remark 6: In fact, a stronger relationship that $\widetilde{W_1^{(i,N)}}$ is equivalent to $\tilde{W}_1^{(i,N)}$ can be proved. This is because that the output symbols combined in the channel degradation process have the same LR. An evidence of this result can be found in [36, Equation (36)], where $\tilde{Z}(\tilde{W}_1^{(i,N)}) = Z(U_1^i | U_1^{1:i-1}, Z^{[N]}) = \tilde{Z}(\widetilde{W_1^{(i,N)}})$. Nevertheless, the degradation relationship is sufficient for this work. Notice that Lemma 12 can be generalized to high level ℓ , with outputs $Z^{[N]}$ replaced by $(Z^{[N]}, X_{1:\ell-1}^{[N]})$.

Illuminated by Lemma 12, we can also symmetrize the new induced channel at level ℓ and show that it is degraded with respect to the randomness-induced channel constructed from \tilde{W}_ℓ . For simplicity, letting $\ell = 1$, the new induced channel at level 1 is $\mathcal{Q}_N(W_1, \mathcal{S}_1, \mathcal{R}_1) : U_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c} \rightarrow Z^{[N]}$, which is symmetrized to $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1) : \tilde{U}_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c} \rightarrow (Z^{[N]}, \tilde{U}_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c} \oplus U_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c})$ in the same fashion as in Lemma 8. Recall that the randomness-induced channel of \tilde{W}_1 defined in [8] can be denoted as $\mathcal{Q}_N(\tilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1) : \tilde{U}_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c} \rightarrow (Z^{[N]}, \tilde{X}_1^{[N]} \oplus X_1^{[N]})$. Note that for the randomness-induced channel $\mathcal{Q}_N(\tilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1)$, set $\mathcal{R}_1 \cup \mathcal{S}_1$ is fed with uniformly random bits, which is different from the shaping-induced channel.

Lemma 13: For an asymmetric channel $W_1 : X_1 \rightarrow Z$ and its symmetrized channel $\tilde{W}_1 : \tilde{X}_1 \rightarrow (Z, \tilde{X}_1 \oplus X_1)$, the symmetrized version of the new induced channel $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1)$ is degraded with respect to the randomness-induced channel $\mathcal{Q}_N(\tilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1)$.

Proof: The proof is similar to that of Lemma 12. For a fixed realization $x_1^{[N]}$ and input $\tilde{u}_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c}$, there are $2^{|\mathcal{S}_1 \cup \mathcal{R}_1|}$ choice of $\tilde{x}_1^{[N]}$ remaining. Since $z^{[N]}$ is only dependent on $x_1^{[N]}$, we can build a middle channel which merges the $2^{|\mathcal{S}_1 \cup \mathcal{R}_1|}$ output symbols of $\mathcal{Q}_N(\tilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1)$ to one output symbol of $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1)$, which means that $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1)$ is degraded with respect to $\mathcal{Q}_N(\tilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1)$. Again, this result can be generalized to higher levels. \square

Finally, we are ready to prove the semantic security of our wiretap coding scheme. For brevity, let $M_\ell F_\ell$ and $\tilde{M}_\ell \tilde{F}_\ell$ denote $U_\ell^{(S_\ell \cup \mathcal{R}_\ell)^c}$ and $\tilde{U}_\ell^{(S_\ell \cup \mathcal{R}_\ell)^c}$, respectively. Recall that M is divided into M_1, \dots, M_r at each level. We express MF and $\tilde{M}\tilde{F}$ as the collection of message and frozen bits on all levels of the new induced channel and the symmetric randomness-induced channel, respectively. We also define $\tilde{M}\tilde{F} \oplus MF$ as the operation $\tilde{M}_\ell \tilde{F}_\ell \oplus M_\ell F_\ell$ from level 1 to level r .

Theorem 4 (Semantic security): For arbitrarily distributed message M , the information leakage $I(M; Z^{[N]})$ of the proposed wiretap lattice code is upper-bounded as

$$I(M; Z^{[N]}) \leq I(\tilde{M}\tilde{F}; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF) \leq rN2^{-N^{\beta'}},$$

where $I(\tilde{M}\tilde{F}; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF)$ is the capacity of the symmetrized channel derived from the non-binary channel $MF \rightarrow Z^{[N]}$ ¹¹.

Proof: By [8, Proposition 16], the channel capacity of the randomness-induced channel $\mathcal{Q}_N(\tilde{W}_1, S_1, \mathcal{R}_1)$ is upper-bounded by $N2^{-N^{\beta'}}$ when partition rule (4) is used. By channel degradation, the channel capacity of the symmetrized new induced channel $\tilde{\mathcal{Q}}_N(W_1, S_1, \mathcal{R}_1)$ can also be upper-bounded by $N2^{-N^{\beta'}}$. Since this result can be generalized to higher level ℓ ($\ell \geq 1$), we obtain $C(\tilde{\mathcal{Q}}_N(W_\ell, S_\ell, \mathcal{R}_\ell)) \leq N2^{-N^{\beta'}}$, which means $I(\tilde{M}_\ell \tilde{F}_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}, \tilde{M}_\ell \tilde{F}_\ell \oplus M_\ell F_\ell) \leq N2^{-N^{\beta'}}$. Similarly to (15), we have

$$\begin{aligned} & I(\tilde{M}\tilde{F}; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF) \\ &= \sum_{\ell=1}^r I(\tilde{M}_\ell \tilde{F}_\ell; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF | \tilde{M}_{1:\ell-1} \tilde{F}_{1:\ell-1}) \\ &= \sum_{\ell=1}^r H(\tilde{M}_\ell \tilde{F}_\ell | \tilde{M}_{1:\ell-1} \tilde{F}_{1:\ell-1}) - H(\tilde{M}_\ell \tilde{F}_\ell | Z^{[N]}, \tilde{M}\tilde{F} \oplus MF, \tilde{M}_{1:\ell-1} \tilde{F}_{1:\ell-1}) \\ &\leq \sum_{\ell=1}^r H(\tilde{M}_\ell \tilde{F}_\ell) - H(\tilde{M}_\ell \tilde{F}_\ell | Z^{[N]}, \tilde{M}\tilde{F} \oplus MF, \tilde{M}_{1:\ell-1} \tilde{F}_{1:\ell-1}) \\ &= \sum_{\ell=1}^r I(\tilde{M}_\ell \tilde{F}_\ell; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF, \tilde{M}_{1:\ell-1} \tilde{F}_{1:\ell-1}) \\ &\stackrel{(a)}{=} \sum_{\ell=1}^r I(\tilde{M}_\ell \tilde{F}_\ell; Z^{[N]}, M_{1:\ell-1} F_{1:\ell-1}, \tilde{M}_\ell \tilde{F}_\ell \oplus M_\ell F_\ell) \\ &\stackrel{(b)}{\leq} \sum_{\ell=1}^r I(\tilde{M}_\ell \tilde{F}_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}, \tilde{M}_\ell \tilde{F}_\ell \oplus M_\ell F_\ell) \\ &\leq rN2^{-N^{\beta'}}, \end{aligned}$$

where equality (a) holds because $Z^{[N]}$ is determined by MFR and $\tilde{M}_\ell \tilde{F}_\ell$ is independent of $\tilde{M}_{\ell+1:r} \tilde{F}_{\ell+1:r} \oplus M_{\ell+1:r} F_{\ell+1:r}$, and inequality (b) holds because adding more variables will not decrease the mutual information.

¹¹The symmetrization of a non-binary channel is similar to that of a binary channel as shown in Lemma 8. When X and \tilde{X} are both non-binary, $X \oplus \tilde{X}$ denotes the result of the exclusive or (xor) operation of the binary expressions of X and \tilde{X} .

Therefore, we have

$$\begin{aligned}
I(\mathbf{M}; \mathbf{Z}^{[N]}) &\leq I(\mathbf{MF}; \mathbf{Z}^{[N]}) \\
&\stackrel{(a)}{\leq} H(\tilde{\mathbf{M}}\tilde{\mathbf{F}} \oplus \mathbf{MF}) - H(\mathbf{MF}) + I(\mathbf{MF}; \mathbf{Z}^{[N]}) \\
&\stackrel{(b)}{=} I(\tilde{\mathbf{M}}\tilde{\mathbf{F}}; \mathbf{Z}^{[N]}, \tilde{\mathbf{M}}\tilde{\mathbf{F}} \oplus \mathbf{MF}) \\
&\leq rN2^{-N^{\beta'}},
\end{aligned}$$

where the equality in (a) holds iff MF is also uniform, and (b) is due to the chain rule. \square

E. Reliability

The reliability analysis in Sect. III-D holds for the wiretap coding without shaping. When shaping is involved, the problematic set \mathcal{D}_ℓ at each level is included in the shaping set \mathcal{S}_ℓ and hence determined by the random mapping $\Phi_{\mathcal{S}_\ell}$. In this subsection, we propose two decoders to achieve reliability for the shaping case. The first one requires a private link between Alice and Bob to share the random mapping $\Phi_{\mathcal{S}_\ell}$ and the second one uses the Markov block coding technique [9] without sharing the random mapping. Note that in [6], the message is simply recovered by decoding the fine lattice Λ_b . When instantiated with a polar lattice, the existence of problematic set \mathcal{D}_ℓ does not permit decoding in this straightforward way. Yet, this is only a limitation of SC decoding, not that of the proposed coding scheme¹².

Decoder 1: If $\Phi_{\mathcal{S}_\ell}$ is secretly shared between Alice and Bob, the bits in \mathcal{D}_ℓ can be recovered by Bob simply by the shared mapping. By Theorem 3, the reliability at each level can be guaranteed by uniformly distributed independent frozen bits and a random mapping $\Phi_{\mathcal{S}_\ell}$ according to $P_{\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}}$ at each level. The decoding rule is given as follows.

- Decoding: The decoder receives $y^{[N]}$ and estimates $\hat{u}_\ell^{[N]}$ based on the previously recovered $x_{1:\ell-1}^{[N]}$ according to the rule

$$\hat{u}_\ell^i = \begin{cases} u_\ell^i, & \text{if } i \in \mathcal{F}_\ell \\ \Phi_i(\hat{u}_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}), & \text{if } i \in \mathcal{S}_\ell \\ \underset{u}{\operatorname{argmax}} P_{\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{Y}^{[N]}}(u | \hat{u}_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}, y^{[N]}), & \text{if } i \in \mathcal{I}_\ell \end{cases}.$$

Note that probability $P_{\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{Y}^{[N]}}(u | \hat{u}_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}, y^{[N]})$ can be calculated by the SC decoding algorithm efficiently, treating \mathbf{Y} and $\mathbf{X}_{1:\ell-1}$ (already decoded by the SC decoder at previous levels) as the outputs of the asymmetric channel. As a result, the expectation of the decoding error probability over the randomized mappings satisfies $E_{\Phi_{\mathcal{S}_\ell}}[P_e(\Phi_{\mathcal{S}_\ell})] = O(2^{-N^{\beta'}})$ for any $\beta' < \beta < 0.5$.

Consequently, by the union bound for multilevel decoding, the expected block error probability of our wiretap coding scheme vanishes as $N \rightarrow \infty$. However, this decoder requires the mapping $\Phi_{\mathcal{S}_\ell}$ is only shared between Alice

¹²In fact, the simulation in [8] showed that the unpolarized set is often empty for reasonable parameters. Even if it is not empty, its proportion is vanishing, and one can do some enumeration, list decoding etc.

and Bob. To share this mapping, we can let Alice and Bob have access to the same source of randomness. This means that we may need a private link between Alice and Bob before the described wiretap coding. Fortunately, the rate of this private link can be made vanishing since the proportion of the shaping bits covered by the mapping $\Phi_{\mathcal{S}_\ell}$ can be significantly reduced.

Recall that the shaping set \mathcal{S}_ℓ is defined by

$$\mathcal{S}_\ell \triangleq \{i \in [N] : Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta} \text{ or } 2^{-N^\beta} < Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{Y}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta}\}. \quad (25)$$

It has been shown in [36] that the shaping bits in the subset $\{i \in [N] : Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}) \leq 2^{-N^\beta}\}$ can be recovered according to the rule

$$u_\ell^i = \underset{u}{\operatorname{argmax}} P_{\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{1:N}}(u | u_\ell^{1:i-1}, x_{1:\ell-1}^{1:N}) \quad \text{if } Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}) \leq 2^{-N^\beta},$$

instead of using the mapping. This modification does not change the result of Theorem 3 and a proof can be found in [40] and [41]. As a result, the mapping only has to cover the unpolarized set

$$d\mathcal{S}_\ell = \{i \in [N] : 2^{-N^\beta} < Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta} \text{ or } 2^{-N^\beta} < Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{Y}^{1:N}, \mathbf{X}_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta}\},$$

whose proportion $\frac{|d\mathcal{S}_\ell|}{N} \rightarrow 0$ as $N \rightarrow \infty$.

Remark 7: When $\Phi_{\mathcal{S}_\ell}$ is shared with Bob, the decoding of Λ_b is equivalent to MMSE lattice decoding proposed in [6]. More precisely, by [15, Lemma 7], SC decoding of an asymmetric channel can be converted to that of its symmetrized channel, which is equivalent to an MMSE-scaled channel for lattice Gaussian shaping [15, Lemma 9].

Decoder 2: Alternatively, one can also use the block Markov coding technique [9] to achieve reliability without sharing $\Phi_{\mathcal{S}_\ell}$. As shown in Fig. 8, the message at ℓ -th level is divided into k_ℓ blocks. Denote by $\Delta\mathcal{S}_\ell$ the bits in unpolarized set $d\mathcal{S}_\ell$. The shaping bits \mathcal{S}_ℓ for each block is further divided into unpolarized bits $\Delta\mathcal{S}_\ell$ and polarized shaping bits $\mathcal{S}_\ell \setminus \Delta\mathcal{S}_\ell$. As mentioned above, only $\Delta\mathcal{S}_\ell$ needs to be covered by mapping and its proportion is vanishing. We can sacrifice some message bits to convey $\Delta\mathcal{S}_\ell$ for the next block without involving significant rate loss. These wasted message bits are denoted by \mathbf{E}_ℓ . For encoding, we start with the last block (Block k_ℓ). Given \mathbf{F}_ℓ , \mathbf{M}_ℓ (no \mathbf{E}_ℓ for the last block) and \mathbf{R}_ℓ , we can obtain $\Delta\mathcal{S}_\ell$ according to $\Phi_{\mathcal{S}_\ell}$. Then we copy $\Delta\mathcal{S}_\ell$ of the last block to the bits \mathbf{E}_ℓ of its previous block and do encoding to get the $\Delta\mathcal{S}_\ell$ of block $k_\ell - 1$. This process ends until we get the $\Delta\mathcal{S}_\ell$ of the first block. This scheme is similar to the one we discussed in Sect. III-D. To achieve reliability, we need a secure code with vanishing rate to convey the bits $\Delta\mathcal{S}_\ell$ of the first block to Bob. See [42] for an example of such codes. To guarantee an insignificant rate loss, k_ℓ is required to be sufficiently large. We may set $k_\ell = O(N^\alpha)$ for some $\alpha > 0$.

We also note that it is easy to satisfy the reliability condition when the message bits are not uniformly distributed. Using some additional shared randomness (which can be public), Alice can generate an uniformly random binary sequence \mathbf{M}'_ℓ which has the same length of \mathbf{M}_ℓ and share it with Bob. Instead of encoding \mathbf{M}_ℓ directly, Alice treats

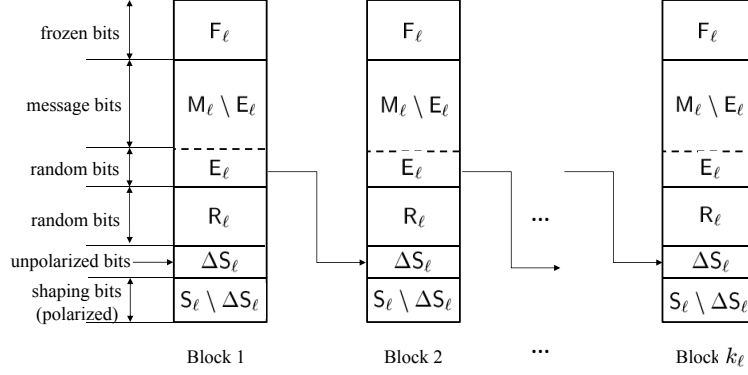


Fig. 8. Markov block coding scheme without sharing the secret mapping.

$M'_\ell \oplus M_\ell$ as the message, which is now uniform. Clearly, $M'_\ell \oplus M_\ell$ can be reliably decoded by both Decoder 1 and Decoder 2. Therefore, M_ℓ can be recovered since Bob knows M'_ℓ .

Now we present the main theorem of the paper.

Theorem 5 (Achieving secrecy capacity of the GWC): Consider a multilevel lattice code constructed from polar codes based on asymmetric channels and lattice Gaussian shaping D_{Λ, σ_s} . Given $\sigma_e^2 > \sigma_b^2$, let $\epsilon_\Lambda(\tilde{\sigma}_e)$ be negligible and set the number of levels $r = O(\log \log N)$ for $N \rightarrow \infty$. Then all strong secrecy rates R satisfying $R < \frac{1}{2} \log \left(\frac{1 + \text{SNR}_b}{1 + \text{SNR}_e} \right)$ are achievable for the Gaussian wiretap channel under semantic security, where SNR_b and SNR_e denote the SNR of the main channel and wiretapper's channel, respectively.

Proof: The reliability condition and the strong secrecy condition are satisfied by Theorem 3 and Lemma 11, respectively. It remains to illustrate that the secrecy rate approaches the secrecy capacity. For some $\epsilon' \rightarrow 0$, we have

$$\begin{aligned}
 \lim_{N \rightarrow \infty} R &= \sum_{\ell=1}^r \lim_{N \rightarrow \infty} \frac{|\mathcal{A}_\ell^{S^c}|}{N} \\
 &= \sum_{\ell=1}^r I(X_\ell; Y | X_1, \dots, X_{\ell-1}) - I(X_\ell; Z | X_1, \dots, X_{\ell-1}) \\
 &\stackrel{(a)}{=} \frac{1}{2} \log \left(\frac{\tilde{\sigma}_e^2}{\tilde{\sigma}_b^2} \right) - \epsilon' \\
 &\stackrel{(b)}{\geq} \frac{1}{2} \log \left(\frac{1 + \text{SNR}_b}{1 + \text{SNR}_e} \right) - \epsilon',
 \end{aligned} \tag{26}$$

where (a) is due to Lemma 10, and (b) is because the signal power $P_s \leq \sigma_s^2$ by Lemma 1¹³, respectively. \square

V. DISCUSSION

We would like to elucidate our coding scheme for the Gaussian wiretap channel in terms of the lattice structure. In Sect. III, we constructed the AWGN-good lattice Λ_b and the secrecy-good lattice Λ_e without considering the power constraint. When the power constraint is taken into consideration, the lattice Gaussian shaping was implemented in

¹³Of course, R cannot exceed the secrecy capacity, so this inequality implies that P_s is very close to σ_s^2 .

Sect. IV. Λ_b and Λ_e were then constructed according to the MMSE-scaled main channel and wiretapper's channel, respectively. We note that these two lattices themselves are generated only if the independent frozen bits on all levels are 0s. Since the independent frozen set of the polar codes at each level is filled with random bits, we actually obtain a coset $\Lambda_b + \chi$ of Λ_b and a coset $\Lambda_e + \chi$ of Λ_e simultaneously, where χ is a uniformly distributed shift. This is because we can not fix the independent frozen bits F_ℓ in our scheme (due to the lack of the proof that the shaping-induced channel is symmetric). By using the lattice Gaussian D_{Λ, σ_s} as our constellation in each lattice dimension, we would obtain D_{Λ^N, σ_s} without coding. Since $\Lambda_e + \chi \subset \Lambda_b + \chi \subset \Lambda^N$, we actually implemented the lattice Gaussian shaping over both $\Lambda_b + \chi$ and $\Lambda_e + \chi$. To summarize our coding scheme, Alice firstly assigns each message $m \in \mathcal{M}$ to a coset $\tilde{\lambda}_m \in \Lambda_b/\Lambda_e$, then randomly sends a point in the coset $\Lambda_e + \chi + \lambda_m$ (λ_m is the coset leader of $\tilde{\lambda}_m$) according to the distribution $D_{\Lambda_e + \chi + \lambda_m, \sigma_s}$ via the shaping operation. This scheme is consistent with the theoretical model proposed in [6].

For semantic security, a symmetrized new induced channel from $\tilde{M}\tilde{F}$ to $(Z^{[N]}, \tilde{M}\tilde{F} \oplus MF)$ was constructed to upper-bound the information leakage. This channel is directly derived from the new induced channel from MF to $Z^{[N]}$. According to Lemma 12, this symmetrized new induced channel is degraded with respect to the symmetric randomness-induced channel from $\tilde{M}\tilde{F}$ to $(Z^{[N]}, \tilde{X}_{1:r}^{[N]} \oplus X_{1:r}^{[N]})$. Moreover, when \tilde{F} is frozen, the randomness-induced channel from \tilde{M} to $(Z^{[N]}, \tilde{X}_{1:r}^{[N]} \oplus X_{1:r}^{[N]})$ corresponds to the Λ_b/Λ_e channel given in Sect. III (with MMSE scaling).

APPENDIX A

PROOF OF LEMMA 3

Proof: It is sufficient to show $I(MF; Z^{[N]}) \leq N \cdot 2^{-N^{\beta'}}$ since $I(M; Z^{[N]}) \leq I(MF; Z^{[N]})$. As has been shown in [8], the induced channel $MF \rightarrow Z^{[N]}$ is symmetric when \mathcal{B} and \mathcal{D} are fed with random bits R . For a symmetric channel, the maximum mutual information is achieved by uniform input distribution. Let $\tilde{U}_{\mathcal{A}}$ and $\tilde{U}_{\mathcal{C}}$ denote independent and uniform versions of M and F and $\tilde{Z}^{[N]}$ be the corresponding channel output. Assuming $i_1 < i_2 < \dots < i_{|\mathcal{A} \cup \mathcal{C}|}$ are the indices in $\mathcal{A} \cup \mathcal{C}$,

$$\begin{aligned}
 I(MF; Z^{[N]}) &\leq I(\tilde{U}^{\mathcal{A}} \tilde{U}^{\mathcal{C}}; \tilde{Z}^{[N]}) \\
 &= \sum_{j=1}^{|\mathcal{A} \cup \mathcal{C}|} I(\tilde{U}^{i_j}; \tilde{Z}^{[N]} | \tilde{U}^{i_1}, \dots, \tilde{U}^{i_{j-1}}) \\
 &= \sum_{j=1}^{|\mathcal{A} \cup \mathcal{C}|} I(\tilde{U}^{i_j}; \tilde{Z}^{[N]}, \tilde{U}^{i_1}, \dots, \tilde{U}^{i_{j-1}}) \\
 &\leq \sum_{j=1}^{|\mathcal{A} \cup \mathcal{C}|} I(\tilde{U}^{i_j}; \tilde{Z}^{[N]}, \tilde{U}^{1:i_j-1}) \\
 &= \sum_{j=1}^{|\mathcal{A} \cup \mathcal{C}|} I(\tilde{W}_N^{(i_j)}) \leq N \cdot 2^{-N^{\beta'}}.
 \end{aligned}$$

□

APPENDIX B

PROOF OF LEMMA 4

Proof: According to the definitions of $\mathcal{G}(\tilde{V})$ and $\mathcal{N}(\tilde{W})$ presented in (4),

$$\begin{aligned}\lim_{N \rightarrow \infty} \frac{|\mathcal{G}(\tilde{V})|}{N} &= \lim_{N \rightarrow \infty} \frac{1}{N} |\{i : \tilde{Z}(\tilde{V}_N^{(i)}) \leq 2^{-N^\beta}\}| = C(\tilde{V}), \\ \lim_{N \rightarrow \infty} \frac{|\mathcal{N}(\tilde{W})|}{N} &= \lim_{N \rightarrow \infty} \frac{1}{N} |\{i : \tilde{Z}(\tilde{W}_N^{(i)}) \geq 1 - 2^{-N^\beta}\}| = 1 - C(\tilde{W}).\end{aligned}$$

Here we define another two sets $\bar{\mathcal{G}}(\tilde{V})$ and $\bar{\mathcal{N}}(\tilde{W})$ as

$$\begin{aligned}\bar{\mathcal{G}}(\tilde{V}) &= \{i : \tilde{Z}(\tilde{V}_N^{(i)}) \geq 1 - 2^{-N^\beta}\}, \\ \bar{\mathcal{N}}(\tilde{W}) &= \{i : \tilde{Z}(\tilde{W}_N^{(i)}) \leq 2^{-N^\beta}\}.\end{aligned}$$

Similarly, we have $\lim_{N \rightarrow \infty} \frac{|\bar{\mathcal{G}}(\tilde{V})|}{N} = 1 - C(\tilde{V})$ and $\lim_{N \rightarrow \infty} \frac{|\bar{\mathcal{N}}(\tilde{W})|}{N} = C(\tilde{W})$. Since \tilde{W} is stochastically degraded with respect to \tilde{V} , $\bar{\mathcal{G}}(\tilde{V})$ and $\bar{\mathcal{N}}(\tilde{W})$ are disjoint with each other [32], then we have

$$\lim_{N \rightarrow \infty} \frac{|\bar{\mathcal{G}}(\tilde{V}) \cup \bar{\mathcal{N}}(\tilde{W})|}{N} = 1 - C(\tilde{V}) + C(\tilde{W}).$$

By the property of polarization, the proportion of the unpolarized part is vanishing as N goes to infinity, i.e.,

$$\begin{aligned}\lim_{N \rightarrow \infty} \frac{|\mathcal{G}(\tilde{V}) \cup \bar{\mathcal{G}}(\tilde{V})|}{N} &= 1, \\ \lim_{N \rightarrow \infty} \frac{|\mathcal{N}(\tilde{W}) \cup \bar{\mathcal{N}}(\tilde{W})|}{N} &= 1,\end{aligned}$$

Finally, we have

$$\lim_{N \rightarrow \infty} \frac{|\mathcal{G}(\tilde{V}) \cap \mathcal{N}(\tilde{W})|}{N} = 1 - \lim_{N \rightarrow \infty} \frac{|\bar{\mathcal{G}}(\tilde{V}) \cup \bar{\mathcal{N}}(\tilde{W})|}{N} = C(\tilde{V}) - C(\tilde{W}).$$

□

APPENDIX C

PROOF OF LEMMA 6

Proof: It is sufficient to demonstrate that channel $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ is degraded with respect to $W'(\mathbf{X}_\ell; \mathbf{Z}|\mathbf{X}_{1:\ell-1})$ and $W'(\mathbf{X}_\ell; \mathbf{Z}|\mathbf{X}_{1:\ell-1})$ is degraded with respect to $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ as well. To see this, we firstly construct a middle channel \hat{W} from $\mathbf{Z} \in \mathcal{V}(\Lambda_r)$ to $\bar{\mathbf{Z}} \in \mathcal{V}(\Lambda_\ell)$. For a specific realization \bar{z} of $\bar{\mathbf{Z}}$, this \hat{W} maps $\bar{z} + [\Lambda_\ell/\Lambda_r]$ to \bar{z} with probability 1, where $[\Lambda_\ell/\Lambda_r]$ represents the set of the coset leaders of the partition Λ_ℓ/Λ_r . Then we obtain channel $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ by concatenating $W'(\mathbf{X}_\ell; \mathbf{Z}|\mathbf{X}_{1:\ell-1})$ and \hat{W} , which means $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ is degraded to $W'(\mathbf{X}_\ell; \mathbf{Z}|\mathbf{X}_{1:\ell-1})$. Similarly, we can also construct a middle channel \check{W} from $\bar{\mathbf{Z}}$ to \mathbf{Z} . For a specific realization \bar{z} of $\bar{\mathbf{Z}}$, this \check{W} maps \bar{z} to $\bar{z} + [\Lambda_\ell/\Lambda_r]$ with probability $\frac{1}{|\Lambda_\ell/\Lambda_r|}$, where $|\Lambda_\ell/\Lambda_r|$ is the order of this partition. This means that $W'(\mathbf{X}_\ell; \mathbf{Z}|\mathbf{X}_{1:\ell-1})$ is also degraded to $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$.

By channel degradation and [33, Lemma 1], letting channel W and W' denote $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ and $W'(X_\ell; Z|X_{1:\ell-1})$ for short, we have

$$\tilde{Z}(W_N^{(i)}) \leq \tilde{Z}(W_N'^{(i)}) \text{ and } \tilde{Z}(W_N^{(i)}) \geq \tilde{Z}(W_N'^{(i)}),$$

$$I(W_N^{(i)}) \leq I(W_N'^{(i)}) \text{ and } I(W_N^{(i)}) \geq I(W_N'^{(i)}),$$

meaning that $\tilde{Z}(W_N^{(i)}) = \tilde{Z}(W_N'^{(i)})$ and $I(W_N^{(i)}) = I(W_N'^{(i)})$. \square

REFERENCES

- [1] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.
- [2] I. Csiszár, "Almost independence and secrecy capacity," *Probl. of Inform. Transmission*, vol. 32, pp. 48–57, 1996.
- [3] S. Leung-Yan-Cheong, "On a special class of wiretap channels," *IEEE Trans. Inf. Theory*, vol. 23, no. 5, pp. 625–627, Sep. 1977.
- [4] S. Goldwasser and S. Micali, "Probabilistic encryption," *J. Comput. Syst. Sci.*, vol. 28, no. 2, pp. 270–299, 1984.
- [5] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Proc. CRYPTO 2012*, ser. Lecture Notes in Computer Science, vol. 7417. Springer-Verlag, 2012, pp. 294–311.
- [6] C. Ling, L. Luzzi, J. Belfiore, and D. Stehle, "Semantically secure lattice codes for the Gaussian wiretap channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6399–6416, Oct. 2014.
- [7] E. Arkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.
- [8] H. Mahdavi and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6428–6443, Oct. 2011.
- [9] E. Şaşoğlu and A. Vardy, "A new polar coding scheme for strong security on wiretap channels," in *Proc. 2013 IEEE Int. Symp. Inform. Theory*, Istanbul, Turkey, July 2013, pp. 1117–1121.
- [10] T. C. Gulcu and A. Barg, "Achieving secrecy capacity of the wiretap channel and broadcast channel with a confidential component," Oct. 2014. [Online]. Available: <http://arxiv.org/abs/1410.3422>
- [11] Y.-P. Wei and S. Ulukus, "Polar coding for the general wiretap channel," in *Proc. 2015 IEEE Inform. Theory Workshop*, Jerusalem, Israel, April 2015, pp. 1–5.
- [12] F. Oggier, P. Solé, and J.-C. Belfiore, "Lattice codes for the wiretap Gaussian channel: Construction and analysis," Mar. 2011. [Online]. Available: <http://arxiv.org/abs/1103.4086>
- [13] A. Ernvall-Hytonen and C. Hollanti, "On the eavesdropper's correct decision in Gaussian and fading wiretap channels using lattice codes," in *Proc. 2011 IEEE Inform. Theory Workshop*, Paraty, Brazil, Oct. 2011, pp. 210–214.
- [14] Y. Yan, C. Ling, and X. Wu, "Polar lattices: Where Arikan meets Forney," in *Proc. 2013 IEEE Int. Symp. Inform. Theory*, Istanbul, Turkey, July 2013, pp. 1292–1296.
- [15] Y. Yan, L. Liu, C. Ling, and X. Wu, "Construction of capacity-achieving lattice codes: Polar lattices," Nov. 2014. [Online]. Available: <http://arxiv.org/abs/1411.0187>
- [16] R. Zamir, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*. Cambridge, UK: Cambridge University Press, 2014.
- [17] E. Abbe and A. Barron, "Polar coding schemes for the AWGN channel," in *Proc. 2011 IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, July 2011.
- [18] A. Joseph and A. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2541–2557, May 2012.
- [19] C. Ling, L. Luzzi, and M. Bloch, "Secret key generation from Gaussian sources using lattice hashing," in *Proc. 2013 IEEE Int. Symp. Inform. Theory*, Istanbul, Turkey, July 2013, pp. 2621–2625.
- [20] M. Hayashi and R. Matsumoto, "Construction of wiretap codes from ordinary channel codes," in *Proc. 2010 IEEE Int. Symp. Inform. Theory*, Austin, USA, June 2010, pp. 2538–2542.
- [21] M. Cheraghchi, F. Didier, and A. Shokrollahi, "Invertible extractors and wiretap protocols," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1254–1274, Feb 2012.

- [22] H. Tyagi and A. Vardy, "Explicit capacity-achieving coding scheme for the Gaussian wiretap channel," in *Proc. 2014 IEEE Int. Symp. Inform. Theory*, Honolulu, USA, June 2014, pp. 956–960.
- [23] M. Bloch and J. Laneman, "Strong secrecy from channel resolvability," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8077–8098, Dec. 2013.
- [24] R. A. Chou, M. R. Bloch, and J. Kliewer, "Low-complexity channel resolvability codes for the symmetric multiple-access channel," in *Proc. 2014 IEEE Inform. Theory Workshop*, Hobart, Australia, Nov. 2014, pp. 466–470.
- [25] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.
- [26] Y. Liang, H. Vincent, and S. Shamai, "Information theoretic security," in *Found. Trends Commun. Inf. Theory*. Norwell, MA, USA: Now Publishers, 2009.
- [27] G. D. Forney Jr., M. Trott, and S.-Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 820–850, May 2000.
- [28] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups*. New York: Springer, 1993.
- [29] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 40, pp. 409–417, Mar. 1994.
- [30] C. Ling and J. Belfiore, "Achieving AWGN channel capacity with lattice Gaussian coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct. 2014.
- [31] E. Arıkan and I. Telatar, "On the rate of channel polarization," in *Proc. 2009 IEEE Int. Symp. Inform. Theory*. Seoul, South Korea: IEEE, June 2009, pp. 1493–1495.
- [32] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2009.
- [33] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.
- [34] R. Fischer, "The modulo-lattice channel: The key feature in precoding schemes," *Int. J. Electron. Commun. (AEÜ)*, vol. 59, no. 4, pp. 244–253, June 2005.
- [35] Y. Yan, L. Liu, and C. Ling, "Polar lattices for strong secrecy over the mod- Λ Gaussian wiretap channel," in *Proc. 2014 IEEE Int. Symp. Inform. Theory*, Honolulu, USA, June 2014, pp. 961–965.
- [36] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013.
- [37] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Commun. Lett.*, vol. 13, no. 7, pp. 519–521, July 2009.
- [38] E. Arıkan, "Source polarization," in *Proc. 2010 IEEE Int. Symp. Inform. Theory*, Austin, USA, June 2010, pp. 899–903.
- [39] M. Mondelli, S. H. Hassani, and R. Urbanke, "How to achieve the capacity of asymmetric channels," Sep. 2014. [Online]. Available: <http://arxiv.org/abs/1103.4086>
- [40] L. Liu and C. Ling, "Polar lattices for lossy compression," Jan. 2015. [Online]. Available: <http://arxiv.org/abs/1501.05683>
- [41] E. E. Gad, Y. Li, J. Kliewer, M. Langberg, A. Jiang, and J. Bruck, "Asymmetric error correction and flash-memory rewriting using polar codes," Oct. 2014. [Online]. Available: <http://arxiv.org/abs/1410.3542>
- [42] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," Mar. 2015. [Online]. Available: <http://arxiv.org/abs/1503.08778>